Complex Variables and Applications

George Sparling

Laboratory of Axiomatics
University of Pittsburgh

Lecture Notes
Monday 4th July 2016

Bounded monotonic real sequences converge

Theorem

Let $X = \{x_n : n \in \mathbb{N}\} \subset \mathbb{R}$ be monotonic.
Then either X is unbounded or X converges.

Proof

We show that X converges if X is bounded and monotonic.
Put $\alpha = \sup(X)$ and $\beta = \inf(X)$. These exist by the axioms of $\mathbb{R}$.
Let a real number $\epsilon > 0$ be given.
Then $\alpha - \epsilon$ is not an upper bound for X and $\beta + \epsilon$ is not a lower bound for X, so $M_\epsilon \in \mathbb{N}$ and $N_\epsilon \in \mathbb{N}$ exist, such that $x_{M_\epsilon} > \alpha - \epsilon$ and $x_{N_\epsilon} < \beta + \epsilon$.
If X is increasing, then $\alpha - \epsilon < x_{M_\epsilon} \leq x_n \leq \alpha$ for $M_\epsilon < n \in \mathbb{N}$.
So for $n > M_\epsilon$, we have $|x_n - \alpha| < \epsilon$. So $\lim_{n\to\infty} x_n = \alpha$.
If instead X is decreasing, then we have $\beta + \epsilon > x_{N_\epsilon} \geq x_n \geq \beta$, for $N_\epsilon < n \in \mathbb{N}$. So for $n > N_\epsilon$, we have $|x_n - \beta| < \epsilon$. By the definition of the limit, we have $\lim_{n\to\infty} x_n = \beta$ and we are done.

A reminder: Bolzano-Weierstrass for reals

Theorem

Let $X = \{x_n : n \in \mathbb{N}\} \subset \mathbb{R}$.
Then either X is unbounded, or X has a convergent subsequence.

Proof

Suppose that X is bounded.
We must prove that X has a convergent subsequence.
It suffices to prove that X has a monotonic subsequence, since any subsequence of a bounded sequence is bounded and since any bounded monotonic sequence is convergent.

- Define $n \in \mathbb{N}$ to be a peak, if $x_m \leq x_n$, for all $m \in \mathbb{N}$, with $m \geq n$.

There are now two cases to consider, according to whether or not there are an infinity of peaks. In each case, we show that X has a bounded monotonic subsequence, so we are done.

The first case: an infinity of peaks

First assume that there are an infinity of peaks.
Taken in order these form a strictly increasing subsequence $n_1 < n_2 < n_3 < \ldots$ of positive integers.
Consider the sub-sequence $Y = \{x_{n_k} : k \in \mathbb{N}\}$ of X.
If the sequence Y is not monotonically decreasing, there are positive integers j and k, with $j < k$, such that $x_{n_k} > x_{n_j}$.
But then $n_k > n_j$, so n_j is not a peak, by the definition of a peak, a contradiction.
So Y is monotonic and bounded, so convergent and we are done.

The second case: a finite number of peaks

Now assume that the number of peaks is finite.
Then there is a positive integer M, such that if $M \leq n \in \mathbb{N}$, then n is not a peak.

- Put $n_1 = M \in \mathbb{N}$.
- Since $n_1 \geq M$, n_1 is not a peak, so $n_2 \in \mathbb{N}$ exists, with $n_2 > n_1 \geq M$, so $n_2 \geq M$, such that $x_{n_2} > x_{n_1}$.
- Since $n_2 \geq M$, n_2 is not a peak, so $n_3 \in \mathbb{N}$ exists, with $n_3 > n_2 \geq M$, so $n_3 \geq M$, such that $x_{n_3} > x_{n_2}$.
- Since $n_3 \geq M$, n_3 is not a peak, so $n_4 \in \mathbb{N}$ exists, with $n_4 > n_3 \geq M$, so $n_4 \geq M$, such that $x_{n_4} > x_{n_3}$.

Continuing this process forever, we recursively generate an infinite strictiy increasing sequence of integers, $\{n_k : k \in \mathbb{N}\}$, such that $x_{n_{k+1}} > x_{n_k}$, for all $k \in \mathbb{N}$. Then the sequence $Y = \{x_{n_k} : k \in \mathbb{N}\}$ is a strictly increasing bounded subsequence of X and so is convergent and we are done.

The meaning of a limit of a complex sequence

If $Z = \{z_n : n \in \mathbb{N}\}$ is a sequence of complex numbers what does it mean to say that the sequence has a limit?

- One answer seems clear: if we write $z_n = x_n + iy_n$, for x_n and y_n real, then we would expect z_n to have a limit if both the real sequences $X = \{x_n : n \in \mathbb{N}\}$ and $Y = \{y_n : n \in \mathbb{N}\}$ have limits and then we would have $\lim_{n\to\infty} z_n = z$, where $\lim_{n\to\infty} x_n = x$, $\lim_{n\to\infty} y_n = y$ and $z = x + iy$.
- We observe that:
  - $\lim_{n\to\infty} x_n = x$ if and only if $\lim_{n\to\infty} |x_n - x| = 0$,
  - $\lim_{n\to\infty} y_n = y$ if and only if $\lim_{n\to\infty} |y_n - y| = 0$.
  - We have both $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$ if and only if $\lim_{n\to\infty} \max(|x_n - x|, |y_n - y|) = 0$.
- Also, using the triangle inequality three times, since $z_n - z = (x_n - x) + i(y_n - y)$, we have:
  - $0 \leq |x_n - x| \leq |z_n - z|$ and $0 \leq |y_n - y| \leq |z_n - z|$,
  - $|z_n - z| \leq |x_n - x| + |y_n - y| \leq 2\max(|x_n - x|, |y_n - y|)$,
  - $0 \leq \max(|x_n - x|, |y_n - y|) \leq |z_n - z| \leq 2\max(|x_n - x|, |y_n - y|)$.

Limits of complex sequences

Applying squeeze, we see that:

- First, if $\lim_{n\to\infty} |z_n - z| = 0$, then $\lim_{n\to\infty} \max(|x_n - x|, |y_n - y|) \to 0$, so $\lim_n \to x_n = x$ and $\lim_n \to y_n = y$.
- Conversely, if $\lim_n \to x_n = x$ and $\lim_n \to y_n = y$, then $\lim_{n\to\infty}(2\max(|x_n - x|, |y_n - y|)) = 0$, so we get $\lim_{n\to\infty} |z_n - z| = 0$.

Accordingly, we make the *definition*:

Definition

If $Z = \{z_n : n \in \mathbb{N}\}$ is a sequence of complex numbers, then we say that Z has limit $z \in \mathbb{C}$, if and only if $\lim_{n\to\infty} |z_n - z| = 0$.

When Z has limit z, we write $\lim_{n\to\infty} z_n = z$, or we say $z_n \to z$, as $n \to \infty$. We also say the sequence Z is convergent with limit z. If the limit does not exist, we say that Z is divergent.

The basic limit theorem

We summarize in a theorem that we have just proved:

- Let $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ be a given complex sequence. For each $n \in \mathbb{N}$, write $z_n = x_n + iy_n$, with x_n and y_n real.
- Put $\Re(Z) = \{x_n : n \in \mathbb{N}\}$ and $\Im(Z) = \{y_n : n \in \mathbb{N}\}$, so $\Re(Z)$ and $\Im(Z)$ are ordinary sequences of real numbers.

Theorem

The complex sequence Z has a limit, if and only if both the real sequences $\Re(Z)$ and $\Im(Z)$ have limits and then we have:

$$\lim_{n\to\infty} z_n = z, \quad \lim_{n\to\infty} x_n = x, \quad \lim_{n\to\infty} y_n = y,$$

$$z = x + iy.$$

Limits are unique

If $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ is a complex sequence with two limits p and q, say, then we have, by definition:

$$|z_n - p| \to 0, \ \ |z_n - q| \to 0, \text{ as } n \to \infty,$$

$$|z_n - p| + |z_n - q| \to 0, \text{ as } n \to \infty.$$

Also, for any $n \in \mathbb{N}$, we have, by the triangle inequality:

$$0 \leq |p - q| = |(p - z_n) + (z_n - q)| \leq |z_n - p| + |z_n - q|.$$

Let $n \to \infty$ in this formula.
Then by squeeze, we get $|p - q| = 0$, so $p = q$.

Cauchy sequences of complex numbers

Definition

If $Z = \{z_n : n \in \mathbb{N}\}$ is a sequence of complex numbers, then we say that Z is Cauchy, if and only if given any real $\epsilon > 0$, there is a real number N_ϵ, such that we have $|z_n - z_m| < \epsilon$, for all positive integers n and m larger than N_ϵ.

Put $x_n = \Re(z_n)$ and $y_n = \Im(z_n)$, so $z_n = x_n + iy_n$, for all $n \in \mathbb{N}$.
Put $\Re(Z) = \{x_n : n \in \mathbb{N}\} \subset \mathbb{R}$ and $\Im(Z) = \{y_n : n \in \mathbb{N}\} \subset \mathbb{R}$.
Using the triangle inequality, as above, we have:

$$0 \leq \max(|x_n - x_m|, |y_n - y_m|) \leq |z_n - z_m| \leq 2\max(|x_n - x_m|, |y_n - y_m|).$$

It follows immediately as above that we have the theorem:

Theorem

The complex sequence Z is Cauchy if and only if both the real sequences $\Re(Z)$ and $\Im(Z)$ are Cauchy.

Cauchy if and only if convergent

Since real sequences are convergent if and only if they are Cauchy, we get the immediate Corollary:

Corollary

A complex sequence $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ is Cauchy if and only if it is convergent.

As with reals we have an ϵ version of convergence:

Theorem

A complex sequence $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ is convergent to the limit $z \in \mathbb{C}$ if and only if given any real $\epsilon > 0$, there is a real N_ϵ, such that, for any $n \in \mathbb{N}$ with $n > N_\epsilon$, we have $|z_n - z| < \epsilon$.

This follows, because the given condition is just the definition of the limit $\lim_{n \to \infty} |z_n - z| = 0$.

Convergent sequences are bounded

Lemma

If a sequence $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ is convergent, then Z is bounded: some positive real M exists, such that $|z_n| \leq M$, for all $n \in \mathbb{N}$.

This follows because, by definition of the limit, if Z has limit z, there is a positive integer N, such that $|z_n - z| < 1$, for all $n \in \mathbb{N}$, such that $n > N$.
Then, for $n > N$, we have, by the triangle inequality:
$|z_n| = |(z_n - z) + z| \leq |z_n - z| + |z| < 1 + |z|$.
So Z is bounded by $M = \max(|z_1|, |z_2|, \ldots, |z|_N, |z| + 1)$ and we are done.

- A corollary is that an unbounded sequence is never convergent.

Convergent sequences multiply

Let $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ and $W = \{w_n : n \in \mathbb{N}\} \subset \mathbb{C}$ be given sequences with limits z and w.
Let Z be bounded by M: $|z_n| \leq M$, for all $n \in \mathbb{N}$.
Put $Y = \{w_n z_n : n \in \mathbb{N}\} \subset \mathbb{C}$.
Then we have, by the triangle inequality:

$$0 \leq |w_n z_n - wz| = |(w_n - w)z_n + w(z_n - z) \leq |(w_n - w)z_n| + |w(z_n - z)|$$

$$= |z_n||w_n - w| + |w(z_n - z)| \leq M|w_n - w| + |z_n - z|.$$

Let $n \to \infty$.

- Then $|w_n - w| \to 0$, so $M|w_n - w| \to 0$. Also $|z_n - z| \to 0$.
- So $M|w_n - w| + |z_n - z| \to 0$.
- So, by squeeze, we get $|w_n z_n - wz| \to 0$.
- So the product sequence Y converges with limit wz.

In particular, if W is a constant sequence, so $w_n = w$, for all $n \in \mathbb{N}$, then it converges with limit w, since $|w_n - w| = 0$, for all $n \in \mathbb{N}$. So the sequence $\{wz_n : n \in \mathbb{N}\}$ converges, with limit wz.

Convergent sequences have convergent sizes

Lemma

If a sequence $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ is convergent, then the real sequence $|Z| = \{|z_n| : n \in \mathbb{N}\} \subset \mathbb{R}$ is convergent with limit $|z|$.

Proof:
By the reverse triangle inequality, for any $n \in \mathbb{N}$, we have:

$$0 \leq ||z_n| - |z|| \leq |z_n - z|.$$

If now the sequence Z has limit z, then $\lim_{n\to\infty} |z_n - z| = 0$, so by squeeze, we have also $\lim_{n\to\infty} ||z_n| - |z|| = 0$, so $\lim_{n\to\infty} |z_n| = |z|$, so the sequence $|Z|$ converges to $|z|$ as required and we are done.
Note that the converse is false, even for real sequences: for example, the sequence $Z = \{(-1)^n : n \in \mathbb{N}\}$ is divergent and yet $|Z| = \{1 : n \in \mathbb{N}\}$ is convergent, with limit 1.

Tails of convergent sequences converge

Let $Z = \{z_n : n \in \mathbb{N}\}$ be a sequence of complex numbers.

- For k a non-negative integer, the k-tail of the sequence is denoted Z_{+k} and is $Z_{+k} = \{z_{n+k} : n \in \mathbb{N}\}$.
- So the k-tail is the subsequence of Z obtained from Z by deleting the first k terms of Z.
- Now let $\lim_{n\to\infty} z_n = z$.
  Then $\lim_{n\to\infty} |z_n - z| = 0$.
- Then for any non-negative integer k, we have also $\lim_{n\to\infty} |z_{n+k} - z| = 0$.
- So Z_{+k} converges, with limit z.
- So the tails of a convergent series converge, all with the same limit as the original series.

We can also state this fact as:

- If we modify a sequence of complex numbers by adding or deleting a finite number of terms, we affect neither its convergence, nor its limit.

Subsequences of convergent sequences converge

Let $Z = \{z_n : n \in \mathbb{N}\}$ is convergent, with limit z.

- Let $Y = \{z_{n_k} : k \in \mathbb{N}\} \subset Z$ be a subsequence of Z. This entails that $n_k \in \mathbb{N}$ strictly increases with $k \in \mathbb{N}$. In particular, by induction, we have: $n_k \geq k$, for all $k \in \mathbb{N}$.
- Then Y has limit z also, since $\{|z_{n_k} - z| : k \in \mathbb{N}\}$ is a real subsequence of the real sequence $\{|z_n - z| : n \in \mathbb{N}\}$. The latter sequence is convergent with limit zero, so the subsequence $\{|z_{n_k} - z| : k \in \mathbb{N}\}$ also converges with limit zero, so Y has limit z, as required.

It is sometimes convenient to use this to prove that a given sequence Z does not converge:

- If Z has a divergent subsequence, then Z diverges.
- If Z has two subsequences X and Y say, each convergent with limits x and y, respectively, then if $x \neq y$, the sequence Z is divergent.

Sequences going to infinity: the Lemma, first part

Let $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$ be a complex sequence.
We say that $\lim_{n\to\infty} z_n = \infty$ if and only if $\lim_{n\to\infty} |z_n| = \infty$, so if and only if, given any positive real M, there is a real number $N(M)$, such that if $n \in \mathbb{N}$ and $n > N(M)$, then $|z_n| > M$.

Lemma

Let $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$. Then $\lim_{n\to\infty} z_n = \infty$ is false if and only if there is a subsequence Y of Z that is bounded.

$\Longleftarrow$

Suppose that Z has a subsequence $Y = \{z_{n_k} : k \in \mathbb{N}\}$ that is bounded. Then there is a positive real M, such that $|z_{n_k}| \leq M$, for all $k \in \mathbb{N}$. Also $n_k \to \infty$ as $k \to \infty$, since $n_k \geq k$, for all $k \in \mathbb{N}$. Then for that real number M, no real $N(M)$ exists such that $|z_n| > M$, for all $n > N(M)$, since for any real $N(M)$, there is a $k \in \mathbb{N}$, such that $k > N(M)$ (Archimedes!), so $n_k > N(M)$ and $|z_{n_k}| > M$ is false. So $\lim_{n\to\infty} z_n = \infty$ is false and we are done.

The Lemma: second part

$\Longrightarrow$
By the contrary of the definition of $\lim_{n\to\infty}|z_n| = \infty$, if $\lim_{n\to\infty} z_n = \infty$ is false, then there is a positive real M, such that for *arbitrarily large* $n \in \mathbb{N}$, we have $|z_n| \le M$.

- First pick $n_1 \in \mathbb{N}$, such that $|z_{n_1}| \le M$.
- Then pick $n_2 \in \mathbb{N}$, such that $|z_{n_2}| \le M$ and $n_2 > n_1$.
- Having picked a strictly increasing sequence $n_1, n_2, \dots, n_k$ of positive integers, such that $|z_{n_j}| \le M$, for all $j \in \mathbb{N}$, with $1 \le j \le k \in \mathbb{N}$, pick $n_{k+1} \in \mathbb{N}$, such that $n_{k+1} > n_k$ and $|z_{n_{k+1}}| \le M$.
- This gives an inductive construction of a subsequence $Y = \{z_{n_k} : k \in \mathbb{N}\}$ of Z, such that $|z_{n_k}| \le M$, for all $k \in \mathbb{N}$, so Y is a bounded subsequence of Z and we are done.

Bolzano-Weierstrass for complex numbers

Theorem

Let $Z = \{z_n : n \in \mathbb{N}\} \subset \mathbb{C}$. Then either $\lim_{n\to\infty} z_n = \infty$, or there is a convergent subsequence of Z.

Proof

Suppose $\lim_{n\to\infty} z_n = \infty$ is false. By the Lemma, there is a bounded subsequence Y of Z. Then the real and imaginary parts of Y, the sequences $\Re(Y)$ and $\Im(Y)$ are bounded real sequences. By the Bolzano-Weierstrass theorem for reals, there is a subsequence X of Y, such that $\Re(X)$ converges. Then $\Im(X)$ is real and bounded, so there is a subsequence W of X, such that $\Im(W)$ converges. But $\Re(W)$ is a subsequence of the convergent sequence $\Re(X)$, so also converges. So $\Re(W)$ and $\Im(W)$ both converge, so W converges. Then W is subsequence of X, X a subsequence of Y, Y a subsequence of Z, so W is a convergent subsequence of Z and we are done.

Sequences of powers

Let z be a fixed complex number. Put $Z = \{z^n : n \in \mathbb{N}\}$.

- If $|z| > 1$, the sequence Z is unbounded and divergent.
- If $|z| < 1$, the sequence Z is bounded in size and converges with limit 0.
- If $|z| = 1$, the sequence Z is bounded in size and divergent unless $z = 1$, in which case Z converges, with limit 1.

Proof:

- If $|z| > 1$, then $|z|^n \to \infty$, as $n \to \infty$, so given any positive real M, there exists $n \in \mathbb{N}$, such that $|z^n| = |z|^n > M$. So the sequence Z is unbounded and therefore divergent, by the Lemma.
- If $|z| < 1$, then $|z|^n \to 0$, as $n \to \infty$, so $\lim_{n\to\infty} |z^n - 0| = \lim_{n\to\infty} |z|^n = 0$, so by definition of the limit, the sequence Z is convergent with limit zero and is therefore bounded also, by the Lemma.

Sequences of powers with $|z| = 1$

The remaining case is the case that $|z| = 1$.

- If $z = 1$, then the sequence is 1 for all n, so converges, with limit 1.

We prove that in all other cases, the sequence diverges.

- If $z^n \to x$, as $n \to \infty$, then $z^{n+1} = z(z^n) \to zx$, since we showed above that multiplying the terms of a convergent sequence by a constant gives a convergent sequence, with limit the constant times the limit of the original sequence. However the sequence $\{z^{n+1} : n \in \mathbb{N}\}$ is a tail of the given sequence Z, so, as we showed above, has the same limit as does Z, namely x, so by the theorem on uniqueness of limits, we get the equation $zx = x$, so $x(1 - z) = 0$. So $x = 0$, or $z = 1$.
- However, since $|z| = 1$, $|z^n| = |z|^n = 1^n = 1 \to 1$, but we showed above that $|z^n| \to |x|$. So $|x| = 1$. So $x \neq 0$.
- So $z = 1$ and we are done.

Summing sequences; the geometric series

Let $Z = \{z_n : n \in \mathbb{N}\}$ is a given sequence of complex numbers.

- For each $n \in \mathbb{N}$, the n-th partial series sum, denoted s_n, of the sequence Z is:

$$s_n = z_1 + z_2 + z_3 + \cdots + z_n.$$

- The limit $S = \lim_{n\to\infty} s_n$ if it exists, is called the series sum of the sequence Z and is written $S = \sum_{n=1}^{\infty} z_n$.

Sometimes it is convenient to start the sum at a term z_0.
Then $s_n = z_0 + z_1 + \cdots + z_{n-1}$ and the infinite sum is written:
$S = \sum_{n=0}^{\infty} z_n = \lim_{n\to\infty} s_n$.
If now $s_n \to S$, then $s_{n+1} \to S$ also, so $s_{n+1} - s_n \to S - S = 0$.
But $s_{n+1} - s_n = z_{n+1}$. So $z_{n+1} \to 0$, so $z_n \to 0$.
We have proved:

- A basic requirement for a series sum to converge is that the sequence $\{z_n : n \in \mathbb{N}\}$ formed by the individual terms of the sequence must converge, with limit zero.

The partial sums of the geometric series

The geometric series is the sum:

$$S(z) = \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \cdots + .$$

Here we have the convention that $z^0 = 1$, for any z.
We have proved above that the sequence $\{z^n : n \in \mathbb{N}\}$ goes to zero if and only if $|z| < 1$.

- So the geometric series *diverges* if $|z| \geq 1$.

Let $z \in \mathbb{C}$ be given and put $s_n = 1 + z + z^2 + \cdots + z^{n-1}$ (n-terms in the sum).

- Then if $z = 1$, we have $s_n = 1 + 1 + 1 + \cdots + 1 = n$.
- Now we claim the following sum formula, valid whenever $z \neq 1$:

$$s_n = 1 + z + z^2 + \cdots + z^{n-1} = \frac{1 - z^n}{1 - z}, \text{ for any } 1 \neq z \in \mathbb{C}.$$

The inductive proof of the sum formula

Proof (by induction), always assuming that $z \neq 1$:

- $s_1 = 1 = \frac{1-z}{1-z}$, so the base case holds.
- If the result is true for $n = k$, we prove it true for $n = k + 1$. We have:

$$s_{k+1} = 1 + z + z^2 + \cdots + z^{k-1} + z^k = s_k + z^k.$$

By the induction hypothesis, we can substitute the formula $\frac{1-z^k}{1-z}$ for s_k, giving:

$$s_{k+1} = s_k + z^k = \frac{1 - z^k}{1 - z} + z^k$$

$$= \frac{1 - z^k + z^k(1 - z)}{1 - z} = \frac{1 - z^k + z^k - z^{k+1}}{1 - z} = \frac{1 - z^{k+1}}{1 - z}.$$

This completes the inductive step, so we are done.

The convergence of the geometric series

We will now prove:

Theorem

The geometric series $S(z) = \sum_{n=1}^{\infty} z^n$ converges if and only if $|z| < 1$ and then the sum is $S(z) = \frac{1}{1-z}$.

Proof:

We have shown above that there is no hope for convergence unless $|z| < 1$. So henceforth, we assume that $|z| < 1$.

Then we have proved above that the partial sums s_n of the geometric series are: $s_n = \frac{1-z^n}{1-z}$, for any $n \in \mathbb{N}$.

Then we have, for any $n \in \mathbb{N}$: $0 \leq \left| s_n - \frac{1}{1-z} \right| = \frac{|z|^n}{|1-z|}$.

But, since $|z| < 1$, the sequence $|z|^n \to 0$, so $\frac{|z|^n}{|1-z|} \to 0$ also.

So by squeeze, we get: $\left| s_n - \frac{1}{1-z} \right| \to 0$.

So $s_n \to \frac{1}{1-z}$ and we are done.

Functions of a complex variable

Our objective here is to develop familiarity with complex functions of one complex variable and how to carry out the basic operations of calculus on such functions.

We have already introduced some basic complex functions:

- z and its powers z^2, z^3, etc.
- Inverse powers z^{-1}, z^{-2}, etc.
- Combining powers of z, with appropriate (complex) constants we construct *polynomials*, such as $z^2 + 4iz - 13$.
- Ratios of polynomials give an important class of functions called *rational functions*, such as $\dfrac{z^2 + 4z + 13}{z^3 - 27i}$.
- Amongst the rational functions there is a distinguished collection *fractional linear*, such as $\dfrac{1 - 2iz}{3z + 4i}$.
- We discuss these in detail and their applications to the *geometry of the sphere*, to basic quantum mechanics, *qubits* and to special relativity, *the intro to Star Trek*).

The composition of fractional linear transformations

Every complex function $f(z)$ is regarded as a map which maps the point z in the complex plane to the image point $w = f(z)$.

Theorem

The composition of two (or more) fractional linear transformations gives another fractional linear transformation.

So let $T(z) = \frac{az+b}{cz+d}$ and $W(z) = \frac{pz+q}{rz+t}$ be arbitrary fractional linear transformations. We compute their composition:

$$(T \circ W)(z) = T(W(z)) = \frac{aW(z)+b}{cW(z)+d} = \frac{a\left(\frac{pz+q}{rz+t}\right)+b}{c\left(\frac{pz+q}{rz+t}\right)+d}$$

$$= \frac{a(pz+q)+b(rz+t)}{c(pz+q)+d(rz+t)} = \frac{(ap+br)z+aq+bt}{(cp+dr)z+cq+dt} = \frac{Az+B}{Cz+D}.$$

Here $A = ap+br, \quad B = aq+bt, \quad C = cp+dr, \quad D = cq+dt$.

Fractional linear transformations and matrices

We can represent fractional linear transformations by a matrix formed by the coefficients:

$$T(z) = \frac{az+b}{cz+d} \quad \leftrightarrow \quad M_T = \begin{vmatrix} a & b \\ c & d \end{vmatrix},$$

$$W(z) = \frac{pz+q}{rz+s} \quad \leftrightarrow \quad M_W = \begin{vmatrix} p & q \\ r & s \end{vmatrix},$$

$$(T \circ W)(z) = \frac{Az+B}{Cz+D} \quad \leftrightarrow \quad M_{T \circ W} = \begin{vmatrix} A & B \\ C & D \end{vmatrix}.$$

Notice now a remarkable fact.
By the calculations on the previous page, we have:

$$M_{T \circ W} = \begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} ap+br & aq+bt \\ cp+dr & cq+dt \end{vmatrix}$$

$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & q \\ r & s \end{vmatrix} = M_T M_W.$$

Fractional linear composition is matrix multiplication

We have proved a nice theorem:

Theorem

If fractional linear transformations T and W have the matrix representations M_T and M_W, respectively, then their composition $T \circ W$ has the matrix representation $M_{T \circ W}$ given by the ordinary matrix product of M_T and M_W:

$$M_{T \circ W} = M_T M_W.$$

Note that for non-constant $T(z)$, we need the rows (a, b) and (c, d) of M_T to be not proportional, so we need the determinant $ad - bc$ to be non-zero. Henceforth for any fractional linear transformation $T(z) = \frac{az+b}{cz+d}$, we shall require that $ad - bc \neq 0$. Then every such transformation is invertible with the inverse transformation represented by the inverse matrix, since $ad - bc \neq 0$ is the condition that the matrix M_T be invertible.

Examples

Recall that the inverse of a two-by-two matrix of non-zero determinant is given as follows:

$$M = \begin{vmatrix} a & b \\ c & d \end{vmatrix}, \quad \mathrm{Adj}(M) = \begin{vmatrix} d & -b \\ -c & d \end{vmatrix},$$

$$M^{-1} = \frac{\mathrm{Adj}(M)}{\det(M)} = \frac{1}{ad - bc} \begin{vmatrix} d & -b \\ -c & a \end{vmatrix}.$$

Let $T(z) = \frac{2z+1}{z+1}$ and $W(z) = \frac{3z+2}{2z+1}$.
We first show that $T \circ T = W$.
Then we compute T^{-1} and the composition $T \circ W$:

$$M_T = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix}, \quad M_{T \circ T} = (M_T)^2 = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} = M_W,$$

$$M_{T^{-1}} = \begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix}, \; M_W = \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}, \; M_{T \circ W} = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} = \begin{vmatrix} 8 & 5 \\ 5 & 3 \end{vmatrix}.$$

The ambiguity in the matrix representation

Consider a fractional linear transformation $T(z) = \frac{az+b}{cz+d}$ and its corresponding matrix $M_T = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

If now we replace (a, b, c, d) by (sa, sb, sc, sd) for any non-zero complex number s, then we get the transformation

$z \to U(z) = \frac{saz+sb}{scz+sd} = \frac{az+b}{cz+d} = T(z)$.

So the transformations $z \to T(z)$ and $z \to U(z)$ coincide.

But the matrix of U is $M_U = \begin{vmatrix} sa & sb \\ sc & sd \end{vmatrix} = sM_T$.

We see that if two matrices U and T are non-zero multiples of each other, then they represent the *same* fractional linear transformation.

We can almost eliminate this ambiguity, by requiring that each matrix representative have determinant 1. If the determinant of the matrix is not one, we have to divide the matrix by the square root of its determinant to make the determinant one.

Using the adjoint matrix for the inverse

For example for the transformation $T(z) = \frac{z+i}{z-i}$, instead of the matrix:

$$M_T = \begin{vmatrix} 1 & i \\ 1 & -i \end{vmatrix}, \quad \det(M_T) = -2i = (1-i)^2,$$

We could use the matrix:

$$M = \frac{1}{1-i}\begin{vmatrix} 1 & i \\ 1 & -i \end{vmatrix} = \frac{1+i}{2}\begin{vmatrix} 1 & i \\ 1 & -i \end{vmatrix} = \begin{vmatrix} \frac{1+i}{2} & \frac{-1+i}{2} \\ \frac{1+i}{2} & \frac{1-i}{2} \end{vmatrix}.$$

Here we have $\det(M) = \frac{1}{4}((1+i)(1-i) - (1+i)(-1+i)) = 1$, as required. Note that there is still an overall sign ambiguity: the matrices M and $-M$ represent the same transformation, and both have determinant one.

An advantage of this ambiguity is that given a matrix M_T, repesenting the fractional linear transformation T, then the adjoint matrix of M_T represents the inverse fractional linear transformation: we do not need to divide by the determinant.

Overview of quantum mechanics

The arena for basic quantum mechanics consists of several parts:

- The Hilbert space of states of a system.
- The collection of observables.
- The notion of measurement.
- The evolution operator.

We consider each in turn.

The Hilbert space of states

- The state space $\mathcal{H}$ is a complex vector space.
  This entails, in particular, that if $|\alpha\rangle$ and $|\beta\rangle$ are states in $\mathcal{H}$, then so is the linear combination $a|\alpha\rangle + b|\beta\rangle$ where a and b are arbitrary *complex* numbers.
  This is called the *Principle of Linear Superposition*.
- There is a notion of size and angle for the space $\mathcal{H}$.

This means that given two states $|\alpha\rangle$ and $|\beta\rangle$ in $\mathcal{H}$, they have a scalar product, denoted $\langle\alpha|\beta\rangle$, which is a complex number, obeying the following rules, valid for any states $|\alpha\rangle$, $|\beta\rangle$ and $|\gamma\rangle$ in $\mathcal{H}$ and for any complex numbers b and c:

- $\langle\beta|\alpha\rangle = \overline{\langle\alpha|\beta\rangle}$.
- $\langle\alpha|\alpha\rangle$ is real and $\langle\alpha|\alpha\rangle \geq 0$; further $\langle\alpha|\alpha\rangle = 0$ iff $|\alpha\rangle = 0$.
- $\langle\alpha|(b|\beta\rangle + c|\gamma\rangle) = b\langle\alpha|\beta\rangle + c\langle\alpha|\gamma\rangle$.

Then the size of a state $|\alpha\rangle$ is $||\alpha\rangle| = \sqrt{\langle\alpha|\alpha\rangle}$.

The physical information of the scalar product

Every non-zero state is a multiple of a normalized state, one of unit size: if $|\alpha\rangle \neq 0$, then $|\gamma\rangle = \dfrac{|\alpha\rangle}{||\alpha\rangle|}$ is normalized: $\langle\gamma|\gamma\rangle = 1$.

States $|\alpha\rangle$ and $|\beta\rangle$ are said to be orthogonal or perpendicular iff $\langle\beta|\alpha\rangle = 0$.

- The triangle inequality holds, for any states $|\alpha\rangle$ and $|\beta\rangle$:

$$||\alpha\rangle + |\beta\rangle| \leq ||\alpha\rangle| + ||\beta\rangle|.$$

- The Cauchy-Schwarz inequality holds for any states $|\alpha\rangle$ and $|\beta\rangle$: $|\langle\beta|\alpha\rangle| \leq ||\alpha\rangle|||\beta\rangle|$.
- In particular, for any normalized states $|\alpha\rangle$ and $|\beta\rangle$, we have $|\langle\beta|\alpha\rangle| \leq 1$.
- Physically, the non-negative real number $|\langle\beta|\alpha\rangle|$ represents the *probability* that a given normalized state $|\alpha\rangle$ is in the normalized state $|\beta\rangle$ (or vice-versa).
- Two non-zero states are physically indistinguishable iff one is a non-zero complex multiple of the other.

Systems with two degrees of freedom

A simple example is the case when the system has two degrees of freedom.
There are then two states $|\uparrow\rangle$ and $|\downarrow\rangle$ and any state $|\psi\rangle$ is a complex linear combination of these states: $|\psi\rangle = p|\uparrow\rangle + q|\uparrow\rangle$ for some complex numbers p and q. We have the inner products:

$$\langle\uparrow|\uparrow\rangle = \langle\downarrow|\downarrow\rangle = 1, \quad \langle\downarrow|\uparrow\rangle = \langle\downarrow|\uparrow\rangle = 0.$$

Then the inner product of the state $|\psi\rangle = p|\uparrow\rangle + q|\uparrow\rangle$ with the state $|\chi\rangle = r|\uparrow\rangle + s|\uparrow\rangle$ is:

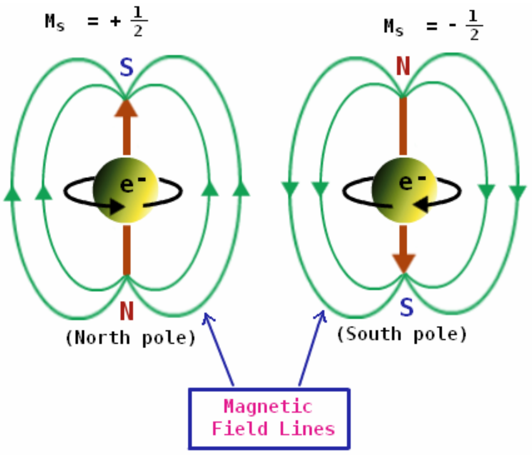
$$\langle\chi|\psi\rangle = (\overline{r}\langle\uparrow| + \overline{s}\langle\downarrow|)(p|\uparrow\rangle + q|\uparrow\rangle) = \overline{r}p + \overline{s}q.$$

In particular we have:

$$\langle\psi|\psi\rangle = \overline{p}p + \overline{q}q = |p|^2 + |q|^2.$$

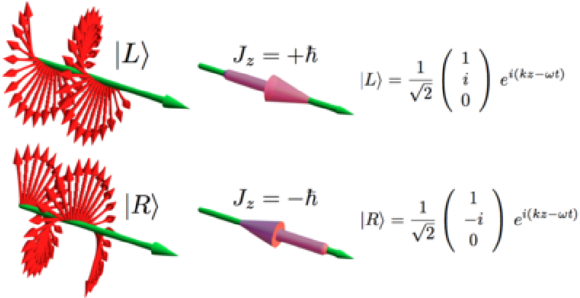
So the state $|\psi\rangle$ is normalized if and only if $|p|^2 + |q|^2 = 1$.

The spinning electron



Two states form the basis for the spin space of an electron: spin up and spin down.

The polarization of the photon



Two states form the basis for the polarization of a photon (or light-wave): counter-clockwise and clockwise.

Two state systems in physics

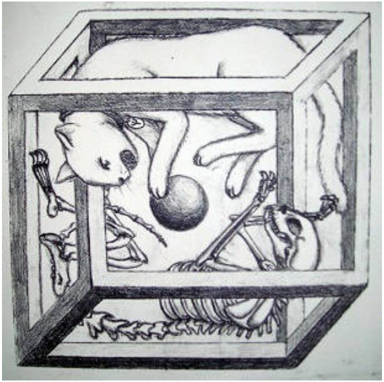
A two-state system can arise in many ways:

- An electron has two basic spin states:
  - $|\uparrow\rangle$ represents an electron spinning up
  - $|\downarrow\rangle$ represents an electron spinning down.

  Later we will see how the general state gives rise to an electron spinning in other directions, so all states are physically realizable and in the end no-one state is preferred over any other.
- A photon has two helicity states:
  - $|\uparrow\rangle$ represents an photon spinning counter-clockwise in its direction of motion.
  - $|\downarrow\rangle$ represents an photon spinning clockwise.

  These are the two basic polarized states; a general state is not necessarily polarized. A polarizer splits the photon into its two different polarized states. In this case there are preferred states, those that are purely polarized, but all states are physically realizable.

Schrödinger's metaphysical cat



Schrödinger's cat has two states dead and alive.
Can it be in a superposition of both?

Schrödinger's famous feline

Schrödinger's cat is a two state space corresponding to a mythical cat in two possible states: $|\uparrow\rangle$ is the state corresponding to the cat being alive and $|\downarrow\rangle$ is the state corresponding to the cat being dead.

- One question is what meaning if any can be attached to complex linear combinations of the two states.
- Another is to ask if the act of measuring the state of the cat forces it to be either dead or alive.

Basic observables

- The quantities $\langle\alpha|$ for $|\alpha\rangle$ in $\mathcal{H}$ are complex linear functionals on $\mathcal{H}$, acting linearly on $\mathcal{H}$ to give a complex number by the formula:

$$\langle\alpha|(|\beta\rangle) = \langle\alpha|\beta\rangle, \text{ for any } |\beta\rangle \text{ in } \mathcal{H}.$$

  They form a Hilbert space, denoted $\mathcal{H}^\dagger$, on an equal footing with the Hilbert space $\mathcal{H}$; following physical practice, we sometimes call $\langle\alpha|$ a bra and $|\alpha\rangle$ a ket.
- Given a state $|\alpha\rangle$ we can create an operator or map, denoted $P_\alpha = |\alpha\rangle\langle\alpha|$ on $\mathcal{H}$, which takes an input state $|\beta\rangle$ to the output state: $P_\alpha|\beta\rangle = |\alpha\rangle\langle\alpha|\beta\rangle$. In particular, if $|\alpha\rangle$ is normalized, the operator acts as the number one on the state $|\alpha\rangle$, i.e. $P_\alpha|\alpha\rangle = |\alpha\rangle$ and annihilates (acts as zero) on any state $|\beta\rangle$ perpendicular to $|\alpha\rangle$, so one for which $\langle\alpha|\beta\rangle = 0$. This is the simplest example of an observable: P_α in some sense recognizes the state $|\alpha\rangle$.

The general definition of observables

- Given the Hilbert space of states, $\mathcal{H}$, an observable A is a linear operator on $\mathcal{H}$: this means that for every state $|\alpha\rangle$ in $\mathcal{H}$, the quantity $A|\alpha\rangle$ is defined and is in $\mathcal{H}$, such that A preserves linear combinations, so for any complex numbers b and c and states $|\beta\rangle$ and $|\gamma\rangle$ in $\mathcal{H}$, we have:

$$A(b|\beta\rangle + c|\gamma\rangle) = bA|\beta\rangle + cA|\gamma\rangle.$$

- We also require that A be compatible with conjugation:

$$\langle\beta|A|\alpha\rangle = \overline{\langle\alpha|A|\beta\rangle}, \text{ for any states } |\alpha\rangle \text{ and } |\beta\rangle \text{ in } \mathcal{H}.$$

Eigen-vectors and the decomposition of an observable

- A basic theorem is that given A, there are states $|\alpha_k\rangle$ that are normalized and perpendicular to each other, such that A may be written (not necessarily uniquely) in the form:

$$A = \sum_{i=1}^{k} \lambda_k P_{\alpha_k} = \sum_{i=1}^{k} \lambda_k |\alpha_k\rangle\langle\alpha_k|, \text{ with each } \lambda_k \text{ real},$$

$$\langle\alpha_k|\alpha_k\rangle = 1, \quad \langle\alpha_j|\alpha_k\rangle = 0, \text{ if } j \neq k.$$

We see that the state $|\alpha_k\rangle$ is an eigen-vector of the observable A of eigen-value λ_k:

$$A|\alpha_k\rangle = \lambda_k|\alpha_k\rangle.$$

We say that A has the value λ_k in the state $|\alpha_k\rangle$.

The Copenhagen Interpretation of Measurement

- If A is an observable, its expectation value $E(A, |\alpha\rangle)$ in the nonzero state $|\alpha\rangle$ is a real number, given by the formula:

$$E(A, |\alpha\rangle) = \frac{\langle\alpha|A|\alpha\rangle}{\langle\alpha|\alpha\rangle}.$$

- In particular if $|\alpha\rangle$ is an eigen-vector of A, with (necessarily real) eigen-value λ, then we have:

$$A|\alpha\rangle = \lambda|\alpha\rangle, \quad E(A, |\alpha\rangle) = \frac{\langle\alpha|A|\alpha\rangle}{\langle\alpha|\alpha\rangle} = \lambda.$$

- The key assumption in the *Copenhagen Interpretation of Quantum Mechanics* has several parts:
  - Any observable can be measured (in principle, although, in practice, it may be difficult to build an appropriate measuring apparatus).
  - If an observable A is measured many times for a quantum system in the state $|\alpha\rangle$, then the average of the measured values of A is given by $E(A, |\alpha\rangle)$.

Wavefront collapse

Perhaps the most controversial idea in the Copenhagen Interpretation is the idea of *wavefront collapse*:

- Suppose the spin of a particle is measured with the result that it is spinning up; then the measurement is repeated almost immediately: one would expect that it is still spinning up.
- In terms of observables this is formulated as:
  - Immediately after the measurement, the state of the system is an eigen-state of the measurement operator with eigen-value corresponding to the measured value.
  - Then if the measurement is repeated, we automatically get the same result for the measurement and will continue to do so for small time intervals as we repeat the measurement.

The up-down observable

In the two-state Hilbert space H, with basis the states $|\uparrow\rangle$ and $\downarrow\rangle$ which are perpendicular and normalized, so that we have the relations $\langle\uparrow|\uparrow\rangle = \langle\downarrow|\downarrow\rangle = 1$ and $\langle\downarrow|\uparrow\rangle = \langle\uparrow|\downarrow\rangle = 0$, the up-down observable is the operator:

$$A = |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|.$$

We see that this obeys the properties:

$$A|\uparrow\rangle = |\uparrow\rangle, \quad A|\downarrow\rangle = -|\downarrow\rangle.$$

Physically we interpret this as:

- If we have a system in the state $|\uparrow\rangle$ and we measure A, we will get the result 1 with certainty, so the particle is definitely (i.e. with probability one) "spinning up" if the system represents the spin of an electron.
- On the other hand if the system in the state $|\downarrow\rangle$ and we measure A, we will get the result -1 with certainty, so the particle is definitely "spinning down".

Mixed states and their measurement

Suppose instead our system is in a mixed normalized state, such as $|\psi\rangle = \frac{3}{5}|\uparrow\rangle + \frac{4}{5}|\downarrow\rangle$. Then we have:

$$A|\psi\rangle = \frac{3}{5}A|\uparrow\rangle + \frac{4}{5}A|\downarrow\rangle = \frac{3}{5}|\uparrow\rangle - \frac{4}{5}|\downarrow\rangle,$$

$$c_\uparrow = \langle\uparrow|A|\psi\rangle = \langle\uparrow|\left(\frac{3}{5}|\uparrow\rangle - \frac{4}{5}|\downarrow\rangle\right) = \frac{3}{5},$$

$$c_\downarrow = \langle\uparrow|A|\psi\rangle = \langle\downarrow|\left(\frac{3}{5}|\uparrow\rangle - \frac{4}{5}|\downarrow\rangle\right) = -\frac{4}{5}.$$

Then the *measurement assumption* is that when the observable A is measured:

- Either the system ends in the eigenstate $|\uparrow\rangle$ with probability $|c_\uparrow|^2 = \frac{9}{25}$
- Or the system ends in the eigen-state $|\downarrow\rangle$ with probability $|c_\downarrow|^2 = \frac{16}{25}$.

The phases of the resulting eigen-states are not specified.

Looking for direction

Throughout the above up-down discussion, we have referred to up and down but no-where is there a discussion of a direct relation to directions in space.
So if we take *any pair* of states $|\alpha\rangle$ and $|\beta\rangle$, which are normalized and perpendicular, then they must specify a direction in space. But which? Let us try to guess. We express the operator which would express "up-downness" for the new states in terms of the old. This operator is:

$$B = |\alpha\rangle\langle\alpha| - |\beta\rangle\langle\beta|.$$

We write: $|\alpha\rangle = p|\uparrow\rangle + q|\downarrow\rangle, \; |\beta\rangle = r|\uparrow\rangle + s|\downarrow\rangle.$
Here p, q, r, s are any complex numbers with $|p|^2 + |q|^2 = |r|^2 + |s|^2 = 1$ and $\bar{r}p + \bar{s}q = 0$.
Then, substituting these expressions into B, we get:

$$B = P|\uparrow\rangle\langle\uparrow| + Q|\uparrow\rangle\langle\downarrow| + \overline{Q}|\downarrow\rangle\langle\uparrow| + R|\downarrow\rangle\langle\downarrow|,$$

$$P = |p|^2 - |r|^2, \; Q = p\bar{q} - r\bar{s}, \; R = |q|^2 - |s|^2.$$

The sphere emerges!

Let us simplify the (p, q, r, s) equations.
Since p and q cannot both be zero, the equation $\bar{r}p + \bar{s}q = 0$ is solved by $r = k\bar{q}$ and $s = -k\bar{p}$, for some complex number k.
Then we need also $1 = |r|^2 + |s|^2 = |k|^2(|p|^2 + |q|^2) = |k|^2$.
So k has unit size, so is a phase, and the only remaining condition is $|p| + |q|^2 = 1$.
Back substituting into (P, Q, R), we get:
$P = |p|^2 - |q|^2, \;\; Q = 2p\bar{q}, \;\; R = |q|^2 - |p|^2 = -P.$
Note that P is real and
$P^2 + |Q|^2 = (|p|^2 - |q|^2)^2 + 4|p|^2|q|^2 = (|p|^2 + |q|^2)^2 = 1.$
Put $Q = x + iy$ and $P = z$, with (x, y, z) real. Then we have:

$$x^2 + y^2 + z^2 = 1,$$

$$x = p\bar{q} + q\bar{p}, \;\; y = -i(p\bar{q} - q\bar{p}), \;\; z = |p|^2 - |q|^2.$$

So the complex numbers (p, q) naturally parametrize a point on a sphere!

Towards Star Trek!

Notice there is another way to describe this parametrization, by dropping the condition that $|p|^2 + |q|^2 = 1$, so we have:

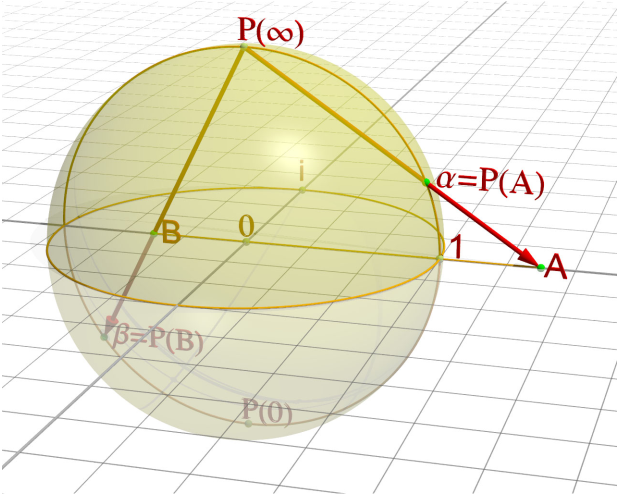
$$x = \frac{p\bar{q} + q\bar{p}}{|p|^2 + |q|^2}, \quad y = \frac{i(q\bar{p} - p\bar{q})}{|p|^2 + |q|^2}, \quad z = \frac{|p|^2 - |q|^2}{|p|^2 + |q|^2}.$$

Here p and q are not both zero. Then the choice $p = 1$ and $q = 0$ gives the point $(0, 0, 1)$, the North Pole, whereas the choice $p = 0$ and $q = 1$ gives the South Pole: $(0, 0, -1)$.

- When $(p, q) = (1, 0)$, so $|\alpha\rangle = |\uparrow\rangle$, we get the standard operator $B = A = |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|$, so we think of the North Pole as representing spin up.
- When $(p, q) = (0, 1)$, so $|\alpha\rangle = |\downarrow\rangle$, we get the negative of the standard operator $B = -A = |\downarrow\rangle\langle\downarrow| - |\uparrow\rangle\langle\uparrow|$, so we think of the South Pole as representing spin down.

So in general we associate to the state $|\alpha\rangle = p|\uparrow\rangle + q|\downarrow\rangle$, the point on the sphere given above.

The stereographic projection of Apollonius



Stereographic projection preserves angles, but not distances.

Stereographic projection

Consider the unit sphere center the origin in three dimensions, so its Cartesian equation is $x^2 + y^2 + z^2 = 1$.

- The North Pole is the point $N = (0, 0, 1)$.
- The South Pole is the point $S = (0, 0, -1)$.
- Let $X = (x, y, z)$ be any point on the sphere. If X is not the North Pole, so if $z \neq 1$, we can join N to X to give a line, $Y = tX + (1-t)Z = (tx, ty, tz + (1-t))$, where t is real.
- This line meets the equatorial plane $z = 0$ at the v,alue of t, such that $tz + 1 - t = 0$, so when $t = (1-z)^{-1}$.
- The point of the plane is then $Y = \left(\frac{x}{1-z}, \frac{y}{1-z}, 0\right)$.

Given X, the point Y is called the stereographic projection of X from the North Pole to the equatorial plane.

We can represent the point Y by a complex number $w = \frac{(x+iy)}{1-z}$, such that the real and imaginary parts of w give the first two co-ordinates of Y, so we have associated a complex number w, to any point X of the sphere, except the North pole.

Solving for the inverse map

Now let us go backwards. Given a complex number $w = u + iv$, representing the point $Y = (u, v, 0)$ of the equatorial plane, we find the point X that stereographically projects to Y, so we want to solve the equations: $\frac{x}{1-z} = u = \Re(w)$, $\frac{y}{1-z} = v = \Im(w)$, such that $x^2 + y^2 + z^2 = 1$. We first find z and $1 - z$ and then $x = \Re(w)(1 - z)$ and $y = \Im(w)(1 - z)$. We have:

$$|w|^2 = u^2 + v^2 = \frac{x^2 + y^2}{(1-z)^2} = \frac{1 - z^2}{(1-z)^2} = \frac{1+z}{1-z},$$

$$|w|^2(1 - z) = 1 + z, \quad z\left(1 + |w|^2\right) = |w|^2 - 1,$$

$$z = \frac{|w|^2 - 1}{|w|^2 + 1}, \quad 1 - z = \frac{|w|^2 + 1 - (|w|^2 - 1)}{|w|^2 + 1} = \frac{2}{|w|^2 + 1}.$$

So we have the required solution:

$$(x, y, z) = \frac{1}{1 + |w|^2}(2\Re(w), 2\Im(w), |w|^2 - 1).$$

The South Pole projection

If, instead, we project the point X of the sphere by joining X to the South Pole S, we get the formula for the projected point Z on the equatorial plane:

$$Z = (r, s, 0), \quad r = \frac{x}{1+z}, \quad s = \frac{y}{1+z}.$$

Now label the point Z by the complex number $v = r - is$. Then we have:

$$v = r - is = \frac{x - iy}{1+z}.$$

So we have, provided X is at neither pole, so that both projections are well-defined:

$$vw = \left(\frac{x - iy}{1+z}\right)\left(\frac{x+iy}{1-z}\right) = \frac{x^2+y^2}{1-z^2} = 1.$$

So v and w are each other's inverses!

The Riemann sphere is a complex manifold!

Note that the co-ordinate w parametrizes all points X on the sphere, except for the North Pole, whereas the co-ordinate v parametrizes all points, except for the South Pole.

- Together all points are covered.
- We say that the sphere is a *complex manifold*: locally it is described by a patch in the complex plane, but we need more than one such patch to cover the whole space.
- Where the patches overlap, the relation between them is given by an holomorphic map.
- Here there are two such patches, one with complex co-ordinate v and one with complex co-ordinate w each ranging over the whole complex plane. These co-ordinates describe the same point when both are non-zero and reciprocals of each other. The function $w \to v = w^{-1}$ is holomorphic, as is the inverse $w = v^{-1}$, so the sphere is a complex manifold. It is called the Riemann sphere.

The inverse and quantum mechanics maps agree!

Now rewrite w as a ratio, so $w = \frac{p}{q}$, where p and q are complex numbers, such that $q \neq 0$. Then we have:

$$(x, y, z) = \frac{1}{1 + \left|\frac{p}{q}\right|^2} \left(\frac{p}{q} + \frac{\overline{p}}{\overline{q}}, -i\left(\frac{p}{q} - \frac{\overline{p}}{\overline{q}}\right), \left|\frac{p}{q}\right|^2 - 1 \right)$$

$$= \frac{1}{|p|^2 + |q|^2} \left(p\overline{q} + q\overline{p}, -i(p\overline{q} - q\overline{p}), |p|^2 - |q|^2 \right)$$

This *exactly agrees* with our formula for the direction associated to the state $|\alpha\rangle = p|\uparrow\rangle + q|\downarrow\rangle$, given above. Note that this formula also encodes the fact that the states $|\alpha\rangle = p|\uparrow\rangle + q|\downarrow\rangle$ and $|\alpha\rangle = tp|\uparrow\rangle + tq|\downarrow\rangle$ for t a non-zero complex number, encode the same information: the formula only depends on the ratio $p : q$ not on p and q individually.

Towards Star Trek: The speed of light

Albert Einstein realized that in order to make concepts of dynamics due to Galileo and Newton compatible with electromagnetic theory, light had to have the property that its *speed was the same in all (inertial) reference frames.*

- Intuitively, if we follows a car along a straight road that is traveling at 100 miles an hour, where our own speed is 60 miles an hour, then the car should be moving at 40 miles an hour relative to us. This is called the Law of Addition of Velocities.
- Einstein realized that as a consequence of Maxwell's equations of electromagnetism that this Law could not be correct and is only a very good approximation at speeds small compared with the velocity of light.

The Relativistic Law of Addition of Velocities

The correct formula deduced by Einstein (and hypothesized earlier by Hendrik Lorentz) is:

- If A, B and C are moving in a straight line in the same direction, such that A is moving at a speed of w relative to B and B is moving at a speed of x relative to C, then the speed v of A relative to C is:

$$v = \frac{w + x}{1 + wx}.$$

  Here we are using units such that the speed of light is one, for example one foot per nanosecond.
- In particular, if $w = 1$, we get $v = \frac{1+x}{1+x} = 1$, *regardless* of the value of x.
- Also if w and x are both small compared with 1, we have v nearly equal to $w + x$, since the factor wx can be ignored.

Hermann Minkowski invents space-time

Hermann Minkowski realized that to understand Einstein's theory properly, space and time had to be unified into a coherent whole.

- In ordinary three dimensional geometry the distance S from the origin to a point (x, y, z) in space is given by the formula of Pythagoras of Samos, $S = \sqrt{x^2 + y^2 + z^2}$ and time plays no direct role.
- Minkowski's space-time, each point (now often called an event) is labelled by four co-ordinates (t, x, y, z) giving the position (x, y, z) of the event and its time t.
- So for example if we are lazy at stay at the origin for all time, we trace the trajectory $(s, 0, 0, 0)$ called our world-line, as s varies.
- If instead we run at speed v in the z-direction, our world-line is $(s, 0, 0, vs)$, as s varies.

Minkowski revamps Pythagoras!

The formula of Minkowski for the squared interval between the origin $O = (0, 0, 0, 0)$ and the point $X = (t, x, y, z)$ is:

$$S^2 = c^2 t^2 - x^2 - y^2 - z^2.$$

Here we are temporarily including c the speed of light in the formula, instead of fixing $c = 1$.

If $t = 0$, we get $-S^2 = x^2 + y^2 + z^2$, recovering the formula of Pythagoras, for the interval between two events occurring with the same t.

We note however that there are now several kinds of intervals:

- $S^2 > 0$ and $t > 0$: X lies to the future of O; a signal can be sent to X from O at less than the speed of light.
- $S^2 > 0$ and $t < 0$: X lies to the past of O; a signal can be sent from X to O at less than the speed of light.
- $S^2 < 0$; X and O are incommunicado, no signal from either can reach the other.

Zero intervals

When $S = 0$, we have the equation $x^2 + y^2 + z^2 = c^2t^2$, or $\rho^2 = c^2t^2$, where we have introduced the ordinary spherical polar co-ordinate $\rho = \sqrt{x^2 + y^2 + z^2} \geq 0$, representing the Pythagorean distance from the origin in space.

$$\rho^2 = c^2t^2, \quad \rho = \pm ct.$$

This defines the *light cone*.

- When $t = 0$, we have just $\rho = 0$ and we are at the origin O.
- When $t > 0$, we are at a distance of ct, exactly the distance that light moves at speed c in time t. This defines the *future light cone* and represents all points X that can be reached by a signal traveling at the spped of light from O.
- When $t < 0$, we are at a distance of $-ct$, exactly the distance that light moves at speed c in time $-t$. This defines the *past light cone* and represents all points X that can be seen by a telescope at the origin.

Parametrizing the null cone

How can we parametrize the null cone?
We already have the answer (where, henceforth, we put $c = 1$)!
Just move the denominator in our parametrization of the sphere into the numerator and make it t, giving the formula:

$$(t, x, y, z) = (|p|^2 + |q|^2, p\bar{q} + q\bar{p}, -i(p\bar{q} - q\bar{p}), |p|^2 - |q|^2).$$

Check: we have:

$$t + z = 2|p|^2, \quad t - z = 2|q|^2, \quad x + iy = 2p\bar{q}, \quad x - iy = 2q\bar{p},$$

$$t^2 - x^2 - y^2 - z^2 = (t - z)(t + z) - (x + iy)(x - iy)$$
$$= 4(|p|^2|q|^2 - p\bar{q}q\bar{p}) = 0.$$

This parametrizes the future null cone; if we want the past null cone, we just replace t by $-t$ in the above formula.

Lorentz Transformations

Following Einstein and Minkowski, consider a transformation from one frame to another which is moving at a uniform speed relative to the original frame.

- Because the speed of light is constant in all frames the null cone is mapped to itself.
- Indeed the space-time squared interval is preserved and the transformation is then a linear transformation on the co-ordinate vector (t, x, y, z) preserving the quantity $t^2 - x^2 - y^2 - z^2$.
- For example, if the transformation preserves the time t, then it preserves $x^2 + y^2 + z^2$, so is an ordinary rotation.
- The full collection of transformations is called the Lorentz group and depends on six parameters, three of which describe rotations and the other three correspond to the transformations to a moving frame, specified by a velocity vector in three dimensions. These are called boosts.

The boost formula

A boost in the z-direction is given by the formulas:

$$x \to x' = x, \quad y \to y' = y,$$

$$t \to t' = at - bz, \quad z \to z' = -bt + az, \quad a^2 - b^2 = 1.$$

Here $b = va$, with v the velocity, and $a > 0$, so we have:

$$a = \frac{1}{\sqrt{1 - v^2}}, \;\; b = \frac{v}{\sqrt{1 - v^2}}.$$

This transformation is linear and we verify that it preserves the squared interval, so that it qualifies as a Lorentz transformation:

$$(t')^2 - (z')^2 = (at - bz)^2 - (-bt + az)^2$$

$$= a^2t^2 - 2abtz + b^2z^2 - b^2t^2 + 2abtz - a^2z^2 = (a^2 - b^2)(t^2 - z^2) = t^2 - z^2,$$

$$(t')^2 - (x')^2 - (y')^2 - (z')^2 = t^2 - x^2 - y^2 - z^2.$$

Note that for z and t small, this reduces approximately to the Galilean formula $t \to t, \;\; z \to z - vt$, v being the speed.

Lorentz transformations for the null cone

How can we rotate the null cone?

- Suppose, for example, we want a rotation in the (x, y) plane, so t and z are fixed.
- Then looking at our (p, q) parametrization, we see that we can achieve this if p and q get multiplied by phases, say $p \to e^{i\alpha}p$ and $q \to e^{i\beta}q$, where α and β are real, since then $|p|^2$ and $|q|^2$ are preserved, so so also are t and z.
- Then, for x and y, putting $\theta = \alpha - \beta$, we get:

$$x+iy = 2p\bar{q} \to x'+iy' = 2p\bar{q}e^{i(\alpha-\beta)} = (x+iy)(\cos(\theta)+i\sin(\theta))$$

$$= x\cos(\theta) - y\sin(\theta) + i(x\sin(\theta) + y\cos(\theta)).$$

$$x \to x' = x\cos(\theta) - y\sin(\theta), y \to y' = x\sin(\theta) + y\cos(\theta).$$

This formula gives the standard rotation in the (x, y)-plane, counter-clockwise through the angle $\theta = \alpha - \beta$.

Boosts for the null cone

How can we boost the null cone?

- Suppose, for example, we want a boost in the (t, z) plane, so x and y are fixed.
- Then looking at our (p, q) parametrization, we see that we can achieve this if p and q get multiplied by real numbers, whose product is one, say $p \to sp$ and $q \to s^{-1}q$, where s is real and non-zero, since then the quantities $p\bar{q}$ and $q\bar{p}$ are preserved, so, so also are x and y.
- Then for t and z, we first define:

$$a = 2^{-1}(s^2 + s^{-2}) > 0, \quad b = 2^{-1}(s^2 - s^{-2}).$$

Note that $a^2 - b^2 = (a + b)(a - b) = s^2(s^{-2}) = 1$.
Then we have:

$$t \to t' = s^2|p|^2 + s^{-2}|q|^2 = 2^{-1}(s^2(t+z) + s^{-2}(t-z)) = at + bz,$$

$$z \to z' = s^2|p|^2 - s^{-2}|q|^2 = 2^{-1}(s^2(t+z) - s^{-2}(t-z)) = bt + az.$$

Relating the parameter s and the boost velocity v

We see that this transformation formula agrees with our earlier formulas and we find the boost velocity is related to the parameter s by the formulas:

$$v = -ba^{-1} = \frac{1-s^4}{1+s^4},$$

$$s^4 = \frac{1-v}{1+v}, \quad s = \left(\frac{1-v}{1+v}\right)^{\frac{1}{4}}.$$

Here $|v| < 1$, but s can have any positive value.

- $s = 1$ corresponds to the identity transformation, so zero speed.
- $s > 1$ corresponds to $v < 0$, going backwards,
- $s < 1$ corresponds to $v > 0$, going forwards.

All Lorentz transformations

We have seen that for prototypical rotations and boosts, we can write formulas for the Lorentz transformation of the null cone in terms of linear action on the (p, q) parameters. We now want to organize all Lorentz transformations. Assemble the space-time co-ordinates in a matrix:

$$X = \left| \begin{array}{cc} t+z & x+iy \\ x-iy & t-z \end{array} \right|.$$

Note that $\det(X) = (t+z)(t-z) - |x+iy|^2 = t^2 - x^2 - y^2 - z^2$.
Also X is hermitian, $X^\dagger = X$, where $X^\dagger$ is the complex conjugate transpose of X.
Note that any 2×2-matrix Y that is equal to its complex conjugate transpose can be written in the form
$\left| \begin{array}{cc} t'+z' & x'+iy' \\ x'-iy' & t'-z' \end{array} \right|$, for unique (t', x', y', z').

All Lorentz transformations

Now the hermitian conjugate of a product of matrices is the product of their conjugates in the opposite order.
Also the hermitian conjugate of the hermitian conjugte of a matrix is itself $(S^\dagger)^\dagger = S$, for any S.
Now suppose we pick any two by two complex matrix
$S = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$ and transform X to $Y = SXS^\dagger$.
Then we see that $Y^\dagger = (S^\dagger)^\dagger X^\dagger S^\dagger = SXS^\dagger = Y$.
So Y gives a new space-time point, (t', x', y', z'), say.
Also the formula for (t', x', y', z'), when written out, is linear in the entries of the original co-ordinate vector (t, x, y, z).
Finally we have:

$$(t')^2-(x')^2-(y')^2-(z')^2 = \det(Y) = \det(SXS^\dagger) = \det(S)\det(X)\det(S^\dagger)$$

$$= \det(X)|\det(S)|^2 = \det(X)|ad-bc|^2 = |ad-bc|^2(t^2-x^2-y^2-z^2).$$

So we have a Lorentz transformation provided $|ad - bc| = 1$.

All Lorentz transformations

Note that replacing S by $e^{iu}S$ for any real u gives the same transformation, since $S^\dagger$ is replaced by $e^{-iu}S^\dagger$.
So after multiplying S by a suitable phase, we may assume without loss of generality that $ad - bc = \det(S) = 1$ for our transformation to be Lorentz.
When X is null, with the (p, q)-parametrization, we get:

$$X = 2\begin{vmatrix} |p|^2 & p\overline{q} \\ q\overline{p} & |q|^2 \end{vmatrix} = 2\begin{vmatrix} p \\ q \end{vmatrix}\begin{vmatrix} \overline{p} & \overline{q} \end{vmatrix}.$$

When we replace $(p, q) \to (p', q')$, where $p' = ap + bq$ and $q' = cp + dq$