8/01/13 Topics in Analysis ①

Some Topics

1. Brouwer's Fixed Point Theorem and related results.

Theorem

If $f: D \to D$ is a continuous map from the closed unit disc $D \subseteq \mathbb{R}^2$ to itself, then $f$ has a fixed point $f(x) = x$.

Related issues: winding numbers, another proof (without complex analysis) of the Fundamental Theorem of Algebra (links to Alg. Top)

2. Approximation by polynomials. Any continuous function $f: [0,1] \to \mathbb{C}$ can be approximated as closely as necessary by polynomials. (Stone Weierstrass)

Links to Linear Analysis. Introduction to interpolation and numerical integration (Chebyshev Polynomials etc)

3. Some number theory topics such as the irrationality of $e$ and $\pi$, the construction of transcendental numbers. Continued fractions and related matters.

4. Baire Category Theorem and weird counterexamples e.g. continuous nowhere-differentiable functions.

Brouwer's Fixed Point Theorem

In the proof of this theorem, we will make crucial use of the Bolzano-Weierstrass property: If $X \subseteq \mathbb{R}^2$ is closed and bounded, and $(x_i)_{i \in \mathbb{N}} \subseteq X$, then we may select a subsequence of the $x_i$ converging to a point of $X$.

The Bolzano-Weierstrass Property is most naturally discussed in the context of a general metric space, where it is known as sequential compactness, the same thing as compactness.

Recall that a metric space is a set $X$ equipped with a distance function $d : X \times X \to \mathbb{R}_{\geq 0}$, such that

i) $d(x, y) = 0 \iff x = y$

ii) $d(x, y) = d(y, x)$

iii) $d(x, z) \leq d(x, y) + d(x, z)$

The crucial example here is $X \subseteq \mathbb{R}^2$, with $d$ the Euclidean metric.

Cauchy Sequences

Let $(x_n)_{n=1}^{\infty} \subseteq X$ be a sequence of elements of $X$.

Formally, for all $\varepsilon > 0$, $\exists N$ such that if $n, m \geq N$, then $d(x_n, x_m) > \varepsilon$. "A sequence that is trying to converge."

Cauchy sequences need not converge, e.g. $X = (0, 1]$, $x_n = \frac{1}{n}$, and also $X = \mathbb{Q}$, $x_n = 3.\underbrace{1415926\ldots}_{n \text{ decimal digits}}$

A metric space is said to be complete if all Cauchy sequences do converge. That is, if $(x_n)_{n=1}^{\infty} \subseteq X$ is a Cauchy sequence, then there is a (unique) point $x$ such that $x_n \to x$.

Formally, this means that for all $\varepsilon > 0$, there exists $N$ such that if $n \geq N$ then $d(x_n, x) < \varepsilon$.

18/01/13

# Topics in Analysis ①

NON - EXAMINABLE !!!

* The real numbers $\mathbb{R}$ can be constructed as the <u>completion</u> of $\mathbb{Q}$

* The real numbers $\mathbb{R}$ are the unique complete metric containing the rationals $\mathbb{Q}$ (with the usual metric) as a dense subset (every real number can be approximated by rationals). How do we construct $\mathbb{R}$? Define $\mathbb{R}$ to be the set of Cauchy sequences in $\mathbb{Q}$, subject to an equivalence relation, namely that $(x_n)_{n=1}^{\infty} \sim (y_n)_{n=1}^{\infty}$ precisely if $x_n - y_n \to 0$. To really set up the real numbers $\mathbb{R}$, one must do many things. Define a distance on $\mathbb{R}$,

$$d\left((x_n)_{n=1}^{\infty}, (y_n)_{n=1}^{\infty}\right) = \cancel{\lim_{n\to\infty} d(x_n, y_n)}\ (d(x_n,y_n))_{n=1}^{\infty}$$

and prove this is well defined. Then one must prove that $\mathbb{R}$ is complete, which gets notationally nasty. Other constructions are Dedekind cuts. However, the completion construction is much more general; one can take any metric space $X$ and form its completion $\tilde{X}$ which is then a complete metric space and contains $X$ as a dense subset. A particular example is the $p$-adics $\mathbb{Q}_p$ ($p$ prime) with respect to the metric defined by $d(x,y) = |x-y|_p$ where

$$|t|_p = p^{-n} \text{ if } t = \frac{p^n a}{b} \text{ with } (a,p) = (b,p) = (a,b) = 1.$$

Note, for example that $2^n \to 0$ in $\mathbb{Q}_2$

<u>Definition</u>

A metric space $X$ is <u>sequentially compact</u> (has the Bolzano-Weierstrass property) if the following holds:

Every sequence in $X$, $(x_n)_{n=1}^{\infty}$ has a convergent subsequence $x_{n_1}, x_{n_2}, \dots$, with $x_{n_i} \to x$.

Theorem

The following are equivalent for a metric space $X$:

1. $X$ is sequentially compact.
2. $X$ is complete and totally-bounded, which means that for every $\varepsilon > 0$, $X$ can be covered by finitely many balls of radius $\varepsilon$.
3. $X$ is compact; every open cover has a finite sub-cover.

Very Sketchy Proof (Sutherland's Metric and Topological Spaces)

$1 \Rightarrow 2$ is relatively easy. If there is a Cauchy sequence which does not converge, then it has no convergent subsequence. If it is not totally bounded, then there is some $\varepsilon > 0$ with an infinite sequence of points $x_n$, each pair $\geq \varepsilon$ apart. This sequence clearly has no convergent subsequence.

21/01/13

# Topics in Analysis ②

## Sequential Compactness / Bolzano-Weierstrass Property

(X compact

(X is sequentially compact) $\Leftrightarrow$ (X is complete and totally bounded)

ii) $\Rightarrow$ i) Suppose that $(x_n)_{n=1}^{\infty} \subseteq X$. Cover X by finitely many balls of radius 1. One of these, $B_1$, contains infinitely many of the $x_n$. Cover this ball by finitely many balls of radius $\frac{1}{2}$. One of these, $B_2$, contains infinitely many $x_n$. Continue with balls of radius $\frac{1}{4}, \frac{1}{8}, \ldots$ Select one point of $(x_n)_{n=1}^{\infty}$ from each ball. This is a Cauchy sequence. Since X is complete, this subsequence converges. $\square$

www.dpmms.cam.ac.uk/~bia23

Last time we talked about $\mathbb{R}$, which was complete (essentially by definition). It follows quite easily that $\mathbb{R}^n$ is complete (if you have a Cauchy sequence, look at it coordinate-wise; each coordinate converges and this gives the limit).

<u>Claim</u>

= seq. compact = has Bolzano-Weierstrass property

$X \subseteq \mathbb{R}^n$ is compact if and only if X is closed and bounded.

<u>Proof</u>

In $\mathbb{R}^n$, totally bounded is the same as bounded (divide into cubes of side length $\varepsilon$). (N.B This uses the fact that $\mathbb{R}^n$ is finite dimensional, and is not true in general. There is a counterexample on the example sheet

In a general complete metric space, a subset is complete if and only if it is closed. If you have a Cauchy sequence in the subset, then it converges in the big space. But then the limit is in the subset. Conversely, if your subset $X$ is not closed, pick a point $x \in Cl(X) = \bar{X}$, but $x \notin X$. Pick a sequence $(x_n)_{n=1}^{\infty} \subseteq X$ and $x_n \in B(x, \frac{1}{n})$. This is then a Cauchy sequence in $X$, which converges to a point not in $X$, namely $x$. $\square$

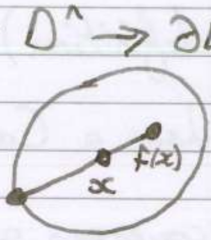
## Brouwer's Fixed Point Theorem

### Theorem

Suppose that $f: D^n \to D^n$ is continuous, where $D^n$ is the closed unit ball in $\mathbb{R}^n$. Then $f$ has a fixed point. This is equivalent to the following

### Proposition

There does not exist a continuous "retraction" $g: D^n \to \partial D^n = S^{n-1}$. That is to say $g(x) = x$ if $x \in \partial D^n$.




### Proposition $\Rightarrow$ Theorem

Suppose that we have a map $f: D^n \to D^n$ without fixed points. Then we define a map $g: D^n \to \partial D^n$ by taking the directed line segment from $f(x)$ to $x$ and extending it to meet $\partial D^n$ (see picture). This map is continuous (example sheet) and preserves the boundary. Hence it is a retraction ※

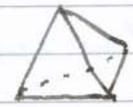
### Theorem $\Rightarrow$ Proposition

If you have a continuous retraction $g: D^n \to \partial D^n$, compose it with a continuous map from $S^{n-1}$ to $S^{n-1}$ without fixed points. (when $n=2$, any non-trivial rotation is fine as is $x \mapsto -x$ for any $n$)

21/01/13

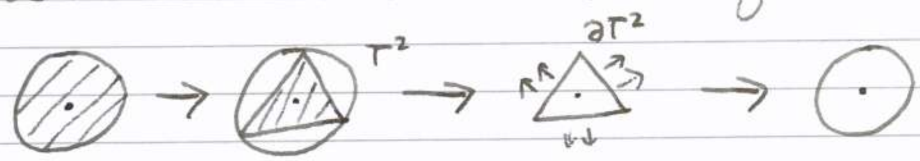
# Topics in Analysis ②

## Equivalent version of Proposition

There does not exist a continuous map $f: T^n \to \partial T^n$ which fixes the boundary. Here, $T^n$ is the $n$-simplex*, which when $n=2$ is just an equilateral triangle. 3-simplex 

## Proof of Equivalence

Suppose that we have a continuous retraction $f: T^n \to \partial T^n$. We can use this to make a continuous retraction of the ball.




The proof of the other direction is identical.

The proof that there is no retraction from $T^n$ to $\partial T^n$ goes via a combinatorial result called Sperner's Lemma.

A triangulation of the simplex $T^n$ is something like this 

A division of the $n$-simplex (when $n=2$, the triangle) into sub-simplices (sub-triangles) which if they meet, do so in a whole face (when $n=2$, the edges, vertices).

A Sperner-colouring of a triangulation is an assignment of colours $\{0, 1, \dots, n\}$ to the vertices of the triangulation such that

i) The $n+1$ vertices of $T^n$ get the colours $0, 1, \dots, n$ once each.

ii) If $v \in \partial T^n$, then $v$ must have ~~ones~~ the same colour as one of the vertices of the $n-1$ dimensional face in which it lies.

iii) If $v$ lies in the interior, its colour can be arbitrary.

## Sperner's Lemma

If you Sperner-colour a triangulation of $T^n$ then there is an elementary sub-simplex whose vertices are labelled $\{0, 1, \dots, n\}$ with each colour appearing exactly once.

not further subdivided

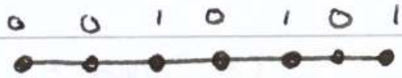
odd number of such simplices

25/01/13

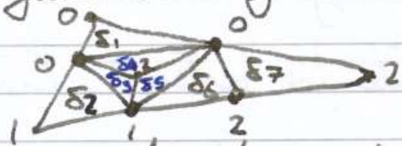
# Topics in Analysis ③

## Proof of Sperner's Lemma

The base case where $n = 1$ is fairly obvious.




The multi-coloured sub-simplices are just the segments with different coloured end-points.




Suppose now that $n \geq 2$. If $\delta$ is an elementary sub-simplex in our triangulation, define $F(\delta)$ to be the number of $(n-1)$-faces of $\delta$, coloured with all of $\{0, \dots, n-1\}$. For example, in the picture, $F(\delta)$ is the number of 0-1 edges of $\delta$, e.g. $F(\delta_7) = 0$, $F(\delta_2) = 2$

Note that $F(\delta) \equiv 0 \ (2)$ unless $\delta$ is multicoloured.

So the number of multicoloured sub-simplices $= \sum_\delta F(\delta) \quad (2)$.

We can evaluate $\sum_\delta F(\delta)$ by looking at the $(n-1)$-faces coloured $\{0, 1, \dots, n-1\}$. Any internal $(n-1)$-face is counted twice, so it makes no contribution mod 2. So the only $(n-1)$-faces that count are on the boundary.

Hence, the number of multicoloured $n$-simplices is congruent to

$\sum_\delta F(\delta) \quad (2) \quad \equiv$ # boundary $(n-1)$ faces coloured $\{0, 1, \dots, n-1\}$

$\equiv 1$ by the inductive-hypothesis, since the colouring of the $\{0, 1, \dots, n-1\}$ boundary face is also a Sperner colouring. $\square$

Back to the Brouwer Fixed Point Theorem:

Suppose we have a continuous map $f : T^n \to \partial T^n$ which fixes the boundary. Triangulate the $n$-simplex into elementary sub-simplices each with diameter $< \varepsilon$ (given some $\varepsilon > 0$).

We Sperner-Colour this triangulation. Assign labels $\{0, 1, \dots, n\}$ arbitrarily to the outside vertices. Let $F_i$ be the $(n-1)$-face opposite the vertex $i$. Given a vertex $v$ in the triangulation, colour it with some colour $i$ such that $f(v) \notin F_i$. This is always possible since $F_1 \cap \dots \cap F_n = \emptyset$. This is easily seen to be a Sperner-colouring (f fixes vertices on the boundary, so we only have the correct colours to choose from.)

By Sperner's Lemma, there is a multicoloured sub-simplex. Suppose that it has vertices $X_0^{(\varepsilon)}, \dots, X_n^{(\varepsilon)}$. Then the distance between each pair of these points is $< \varepsilon$, and $f(X_i^{(\varepsilon)}) \notin F_i$, since $X_i^{(\varepsilon)}$ has colour $i$.

Take a sequence of values of $\varepsilon$ tending to 0, e.g. $1, \frac{1}{2}, \frac{1}{4}, \dots$

Look at the sequence $X_0^{(1)}, X_0^{(\frac{1}{2})}, X_0^{(\frac{1}{4})}, \dots$ all of which lie in the closed bounded set $\Gamma^n$. By Bolzano-Weierstrass, there is a convergent subsequence $X_0^{(\varepsilon_1)}, X_0^{(\varepsilon_2)}, \dots \to x$.

Since $\varepsilon_j \to 0$, we also have $X_1^{(\varepsilon_1)}, X_1^{(\varepsilon_2)}, \dots \to x$, and similarly for $X_2^{(\cdot)}, \dots, X_n^{(\cdot)}$. Now, we coloured so that $f(X_0^{(\varepsilon)}) \notin F_0$. Hence $f(X_0^{(\varepsilon)}) \in F_1 \cup \dots \cup F_n$. By continuity of $f$ and closedness of $F_1 \cup \dots \cup F_n$, we have $f(x) \in F_1 \cup \dots \cup F_n$. $f(x) \in F_i$, some $i$.

But $f(X_i^{(\varepsilon)}) \to f(x)$ $\times\!\!\times$ —

# Topics in Analysis 3

## Winding Numbers and Degree

Recall that if $f: X \to Y$ is a map between metric spaces, then $f$ is continuous at a point $x$, if for all $\varepsilon > 0$, there is a $\delta > 0$ such that $d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \varepsilon$

We say that $f$ is uniformly continuous if the same $\delta$ works for every $x$. For example:

$f: \mathbb{R} \to \mathbb{R}, \; x \mapsto x^2$

is not uniformly continuous.

**Sperner's lemma and Brouwer's fixed point theorem**

The purpose of this note is to correct the oversight made in lectures in connection with this deduction. Best to draw coloured triangles when trying to work through this!

Recall that Sperner's lemma is the following.

**Theorem 1.** *Suppose you triangulate the $n$-simplex $T^n$. Colour the vertices of the triangulation with colours $0, 1, \dots, n$ so that the $n+1$ extremal vertices get each colour once, a vertex in an $(n-1)$-face gets the same colour as one of its $n$ extremal vertices, and all other vertices are coloured arbitrarily (this is called a Sperner colouring). Then there is an elementary subsimplex (minimal simplex in the triangulation) which is multicoloured.*

Recall also that we reduced Brouwer to the following statement:

**Theorem 2.** *There is no continuous retraction $f : T^n \to \partial T^n$, that is to say a continuous map which preserves the boundary.*

Let's show how this follows from Sperner's lemma (you are only really supposed to worry about the case $n = 2$).

Let $\varepsilon > 0$ be arbitrary. Then $T^n$ may be triangulated in such a way that each elementary subsimplex has diameter $< \varepsilon$. This is obvious by picture when $n = 2$, and the case $n \geqslant 3$ is on the example sheet. We're going to make a Sperner colouring of this triangulation. First of all, colour the $n+1$ extreme vertices of $T^n$ arbitrarily with each colour $0, 1, \dots, n$ being used once.

Now (and here we depart from lectures) I claim that there are closed sets $V_0, \dots, V_n \subset \partial T^n$ with the following properties:

(i) $V_0 \cup V_1 \cup \dots \cup V_n = \partial T^n$;

(ii) If $F_i$ is the $(n-1)$-face opposite the extreme vertex of $T^n$ which is coloured $i$, then $F_i \cap V_i = \emptyset$.

In lectures I tried to take $V_i$ to be the union of all faces other than $F_i$, but this won't work as (ii) is violated. I'll leave you to convince yourself that such $V_i$ exist. When $n = 2$, so that $T^2$ is just a triangle labelled with $0, 1, 2$, you can take $V_i$ to be the vertex $i$ together with the 90% nearest $i$ of the two edges containing $i$ (but not 100% as I tried to do in lectures!).

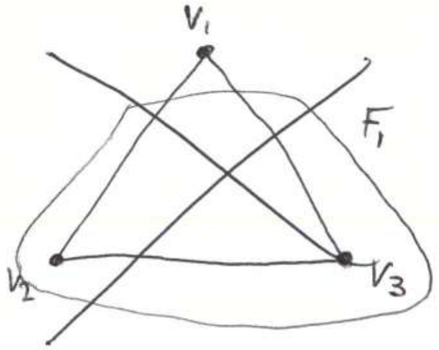
Now colour the vertices of the triangulation in such a way that if $v$ is coloured with $i$ then $f(v) \in V_i$. By property (i) this is possible. Also, if $v$ lies in the face $F_i$ opposite $i$, then by (ii) we see that $v$ is not coloured with $i$. So this is a Sperner colouring.

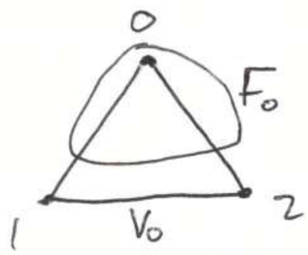
Applying Sperner's lemma, there is a simplex with vertices $x_0^{(\varepsilon)}, \dots, x_n^{(\varepsilon)}$, with $x_i^{(\varepsilon)}$ coloured $i$.

Take a sequence of $\varepsilon$'s tending to 0, say $1, \frac{1}{2}, \frac{1}{4}, \ldots$. By Bolzano-Weierstrass there is a subsequence $x_0^{(\varepsilon_1)}, x_0^{(\varepsilon_2)}, \ldots$ tending to some point $x$. Each sequence $x_j^{(\varepsilon_1)}, x_j^{(\varepsilon_2)}, \ldots$ then tends to $x$ automatically.

Now $x_0^{(\varepsilon_i)}$ is coloured 0, and so $f(x_0^{(\varepsilon_i)}) \in V_0$. Since $V_0$ is closed and $f$ is continuous, we may take limits as $i \to \infty$ to conclude that $f(x) \in V_0$. Similarly $f(x) \in V_1, \ldots, V_n$. But $f(x)$ lies on the boundary $\partial T^n$, and so it must certainly lie in some face $F_i$ and hence not in $V_i$ by property (2).

BJG



25/01/13 Topics in Analysis ④

## Winding Numbers

1. Suppose that $X$ is a (sequentially) compact metric space and that $f: X \to Y$ is a continuous function. Then $f$ is uniformly continuous.

**Sketch Proof**

If not, $\exists \varepsilon > 0$ such that no $\delta$ works. In particular, for each $n$ there are $x_n, x_n'$ such that $d(x_n, x_n') \leq \frac{1}{n}$ but $d(f(x_n), f(x_n'))\geq$

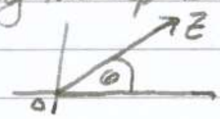
Since $X$ is sequentially compact, there is a subsequence $x_{n_i} \to x$. But then $x_{n_i}' \to x$ as well. Hence by continuity, $f(x_{n_i}), f(x_{n_i}') \to f(x$ $\#$

2. With the same hypothesis, the image of an (sequentially) compact metric space under a continuous map is also (sequentially) compact.

**Sketch Proof**

Consider a sequence $(f(x_n))_{n=1}^{\infty}$ in $f(X)$. Then $(x_n)_{n=1}^{\infty} \subset X$ has a convergent subsequence $x_{n_i} \to x$. But then $f(x_{n_i}) \to f(x)$ by continuity, thus $(f(x_n))_{n=1}^{\infty}$ has a convergent subsequence. □

We will be thinking about paths in $\mathbb{R}^2$, which we will frequently identify with $\mathbb{C}$. Given $z \in \mathbb{C} \setminus \{0\}$, a real number $\theta$ is called an argument of $z$ if $z = |z|e^{i\theta}$. Note that $\theta$ is not uniquely defined; the allowable choices of $\theta$ differ by multiples of $2\pi$.

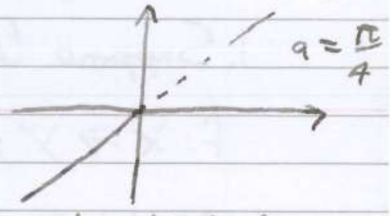



**Lemma**

Consider any domain in $\mathbb{C}$ of the form $X_a = \{z = re^{i\theta} \mid a-\pi < \theta < a+\pi,\ r \neq 0$

Then there is a continuous choice of argument on $X_a$.

We call this $\arg_a$. Given $z$, define $\arg_a(z)$ to be the unique choice of argument in the range $(a-\pi, a+\pi)$.




Brief Justification that $\arg_a$ is continous:

Suppose that $z = re^{i\theta}$, $z' = se^{i\varphi}$. We need to show that if $z, z'$ are closed, then so are $\theta, \varphi$. We have

$$|z-z'|^2 = |re^{i\theta} - se^{i\varphi}|^2 = (r-s)^2 + 2rs(1-\cos(\theta-\varphi))$$

Hence if $z \approx z'$, then $\cos(\theta-\varphi) \approx 1$, but since $a-\pi < \theta, \varphi < a+\pi$ and this implies that $\theta \approx \varphi$. $\square$

Corollary

Suppose that $f : [0,1] \to X_a \subset \mathbb{C} \setminus \{0\}$ is a continuous path. Then there is a continuous choice of argument. Indeed, consider $\arg_a \circ f$. This is continuous since it is a composition of continuous functions.

Proposition

Let $f : [0,1] \to \mathbb{C} \setminus \{0\}$ be a continuous path. Then there is a continuous choice of $\arg(f(t))$.

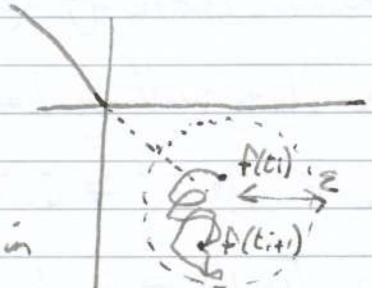
Proof

The idea is to split into subpaths, each covered by the Corollary. Since $[0,1]$ is (sequentially) compact, so is $f([0,1])$. Hence, in particular, $f([0,1])$ is closed. Since $0 \notin f([0,1])$, there is some $\varepsilon > 0$ such that $f([0,1])$ avoids $B_\varepsilon(0)$ i.e. $|f(t)| \geq \varepsilon$ for all $t \in [0,1]$.

By uniform continuity of $f$, we can split $[0,1]$ into sub-intervals

$0 = t_0 < t_1 < \ldots < t_k = 1$

with $t_{i+1} - t_i \leq \delta$, where $\delta$ is chosen such that $|x - x'| \leq \delta$ $\Rightarrow |f(x) - f(x')| \leq \frac{\varepsilon}{2}$.




Looking at the picture, for $t$ in the range $[t_i, t_{i+1}]$, $f(t)$ always lies in the domain $X_a$, where $a$ is some choice of $\arg(f(t_i))$. Hence there is a continuous choice of argument $\arg(f(t))$ for $t \in [t_i, t_{i+1}]$. Splicing these together, adjusting by multiples of $2\pi$ to make the endpoints match, we get a continous choice of $\arg(f(t))$. $\square$

Suppose we now have two continous choices of $\arg(f(t))$, $\theta_1(t)$ and $\theta_2(t)$. Then $\theta_1(t) - \theta_2(t) \in 2\pi\mathbb{Z}$ and this function is continous. By the intermediate value theorem $\theta_1(t) - \theta_2(t)$ is constant. So $\arg(f(t))$ is unique up to a constant multiple of $2\pi$.

Corollary

If $f : [0,1] \to \mathbb{C} \setminus \{0\}$ is continous, then $\frac{1}{2\pi}(\arg(f(0)) - \arg(f(1)))$ does not depend on the choice of arg. If $f$ is a closed path, that is, $f(0) = f(1)$, this quantity is an integer. We write this $w(f, 0)$ and call it the winding number or degree of $f$ about $0$.

Observation

$w(fg, 0) = w(f, 0) + w(g, 0)$. Indeed, if $f(t) = |f(t)| e^{i\arg(f(t))}$ and similarly for $g$, then

$f(t)g(t) = |f(t)g(t)|e^{i(\arg(f(t)) + \arg(g(t)))}$, thus $\arg(f(t)) + \arg(g(t))$ is a continous choice of $\arg(fg)$. (Here, $f, g : [0,1] \to \mathbb{C}\setminus\{0\}$ )

Lemma ('Dog Walking Lemma')

Suppose $f : [0,1] \to \mathbb{C}\setminus\{0\}$ is a continous path with $f(0) = f(1)$.
Let $g : [0,1] \to \mathbb{C}$ be another continous path with $g(0) = g(1)$.
Suppose that $|g(t)| < |f(t)|$ for all $t$.
Then $w(f+g, 0) = w(f, 0)$.
($f$ : path of walker about the lamp-post. $g$ : displacement of the dog from the walker. Assume : the leash is short so the dog cannot touch the lampport)

Proof

Note that $w(f+g, 0) = w(f, 0) + w(1 + g/f, 0)$.
However, $(1 + g/f)(t)$ always lies in the domain $|z - 1| < 1$

and hence in $X_0$. Hence $\arg_0(1 + \frac{g}{f})$ is a continous choice of argument taking values in $-\pi < \theta < \pi$.
Hence, since $\arg_0(1 + g/f)(1)$ and $\arg_0(1 + g/f)(0)$ differ by $2\pi\mathbb{Z}$, they must be equal. So $w(1 + g/f, 0) = 0$

28/01/13 Topics in Analysis 5

## Dog Walking Lemma

If $f : [0,1] \to \mathbb{C} \setminus \{0\}$ has $f(0) = f(1)$, and $\exists\, g : [0,1] \to \mathbb{C}$ with $g(0) = g(1)$, $f, g$ continuous and $|g(t)| < |f(t)|$, then $w(f+g, 0) = w(f, 0)$ (Very Reminiscent of Rouché's Theorem)

## Homotopy Invariance of Winding Number

Suppose we have two closed paths $f_0, f_1 : [0,1] \to \mathbb{C} \setminus \{0\}$. Then we say that $f_0$ and $f_1$ are homotopic, if there is a continuous map ("Continuous movements of a closed path preserve the winding number.")

$F : [0,1] \times [0,1] \to \mathbb{C} \setminus \{0\}$ such that

i) $F(0,t) = f_0(t) \quad \forall t$

ii) $F(1,t) = f_1(t) \quad \forall t$

iii) $F(s,0) = F(s,1)$

## Idea

$F(s,t)$, for fixed $s$, is a closed path, which we could write $f_s(t)$. The $f_s$, $s \in [0,1]$, are a continuous family of closed paths, interpolating between $f_0$ and $f_1$.

## Theorem

Suppose that $f_0, f_1 : [0,1] \to \mathbb{C} \setminus \{0\}$ are closed paths which are homotopic. Then $w(f_0, 0) = w(f_1, 0)$

## Proof

The basic idea is to take a suitable subdivision $0 = t_0 < t_1 < \ldots < t_k = 1$

and check that $w(f_0, 0) = w(f_{t_1}, 0) = \dots = w(f_{t_k}, 0) = w(f_1, 0)$

Note that $[0,1] \times [0,1]$ is sequentially compact, and hence closed. Hence, there is some $\varepsilon > 0$ such that $|F(s,t)| \geq \varepsilon$ for all $s,t \in [0,1] \times [0,1]$. But $F$ is continuous, and hence uniformly continuous, and so $\exists \delta > 0$ such that if $d((s,t),(s',t')) \leq \delta$ then $|F(s,t) - F(s',t')| < \varepsilon$. Suppose now that $|s-s'| \leq \delta$. Then $|f_s(t)| \geq \varepsilon$ $(f_s(t) = F(s,t))$ and $|f_{s'}(t) - f_s(t)| < \varepsilon$ since $d((s,t),(s',t')) \leq \delta$.

Applying the Dog-Walking Lemma, $f = f_s$, $g = f_{s'} - f_s$, we get $w(f_{s'}, 0) = w(f+g, 0) = w(f, 0) = w(f_s, 0)$

Now take a subdivision $0 = s_0 < \dots < s_k = 1$ with $|s_{i+1} - s_i| \leq \delta$

Then $w(f_0, 0) = \dots = w(f_1, 0)$ $\square$

<u>Proposition</u>

Write $D$ for the closed unit disc in $\mathbb{R}^2 = \mathbb{C}$. Suppose that $g : D \to \mathbb{C}$ is a continuous function such that $g \neq 0$ on $\partial D$, and the winding number of the path $t \mapsto g(e^{2\pi i t})$ is not zero. Then there is some $z \in D$ such that $g(z) = 0$

<u>Proof</u>

Suppose not. Consider the map $F : [0,1] \times [0,1] \to \mathbb{C} \setminus \{0\}$ defined by $F(s,t) = g(s e^{2\pi i t})$. This is fairly clearly continuous. It gives us a homotopy between $F(1,t) = g(e^{2\pi i t})$ and

and the constant path $t \mapsto g(0)$. This is a contradiction, since by assumption the first path has non-zero winding number, but the winding number of a constant map is $0$. $\square$

* In the world of complex analysis, the winding number of $g(e^{2\pi i t})$ is actually the number of zeroes of $g$ inside $D$ (counted with multiplicity)
* In Real Analysis this is not true.

Another Proof of Brouwer's Fixed Point Theorem in $\mathbb{R}^2$

It suffices to show that there is no continuous retraction ~~from~~ $g: D \to \partial D$, that is to say that there is no continuous map which fixes $\partial D$. However, such a map $g$ has no zeroes in $D$ and $g(e^{2\pi i t}) = e^{2\pi i t}$ which has winding number 1, because $2\pi t$ is a continuous choice of $\arg(e^{2\pi i t})$, hence $\frac{1}{2\pi}(\arg(e^{2\pi i 1}) - \arg(e^{2\pi i 0})) = 1$.
But this is the winding number of $g(e^{2\pi i t})$, contradicting the proposition $\square$

Proof of the Fundamental Theorem of Algebra

Every non-constant polynomial over $\mathbb{C}$ has a root.

Let $p$ be a polynomial $p(z) = a_n z^n + \dots + a_1 z + a_0$, $a_n \neq 0$, $n \geq 1$. Let $R$ be very big (how big will come later).

We look at the following closed paths from $[0,1] \to \mathbb{C}$.

$u(t) = p(Re^{2\pi i t})$ $\qquad$ $v(t) = a_n(Re^{2\pi i t})^n$

$w(t) = u(t) - v(t)$

Suppose that $p$ has no zero, and suppose that $|w(t)| < |v(t)|$

Then the Dog Walking Lemma implies that ~~the winding number of u~~ $w(u,0) = w(v,0)$.

However, ~~the winding number~~ $w(v,0) = n$ since $2\pi t n$ is a continuous choice of $\arg(v(t))$. But then, by the earlier proposition about functions on $D$, (generalised to functions on the disc of radius $R$), $p$ does have a root $z$ with $|z| \leq R$, ※.

It remains to be shown that if $R$ is sufficiently large, then $|w(t)| < |v(t)|$

But $|w(t)| \leq (|a_0| + \dots + |a_n|) R^{n-1}$ if $R > 1$, by the Triangle Inequality

Hence everything works if $R > \frac{|a_0| + \dots + |a_{n-1}|}{|a_n|}$, since $|v(t)| = |a_n| R^n$.

30/01/13 Topics in Analysis ⑥

## Chapter 2: Approximation of Continuous Functions by Polynomials

Let $f:[a,b]\to\mathbb{R}$ be continuous. How closely can $f$ be approximated by polynomials? There is no loss of generality in considering $[0,1]$. What do we mean by approximate? There are many ways to measure the "distance" between functions and hence to say how good our approximations are. e.g. $\|f-g\|_1 = \int_0^1 |f(x)-g(x)|\,dx$ or $\|f-g\|_2 = \left(\int_0^1 |f(x)-g(x)|^2 dx\right)^{\frac{1}{2}}$. We will be concerned exclusively with the sup-norm of $L^\infty$, defined by $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$. If $f$ is continuous, then this sup is also a maximum (recalling that a continuous function on a closed, bounded interval such as $[0,1]$ is bounded and attains its bounds. Indeed, $f([0,1])$ is sequentially compact, so is closed and bounded). For brevity, write $\|f\| = \|f\|_\infty$. Recall that this gives a distance $d(f,g) = \|f-g\|$. This gives the space $C[0,1]$ of continuous real-valued functions the structure of a metric space, which turns out to be complete.

We will prove the following facts:

1. Every $f\in C[0,1]$ can be uniformly approximated by polynomials. That is, for every $\varepsilon>0$, $\exists\, p$, a polynomial such that $\|f-p\|<\varepsilon$. Put another way, the polynomials are dense in $C[0,1]$. This is a special case of the <u>Stone-Weierstrass Approximation Theorem</u>.
2. Among polynomials of degree at most $n$, there is a unique best

approximation to $f$, i.e. a polynomial $p_*$ such that $\|f - p_*\| < \|f - p\|$ for all $p \in \mathcal{P}_n \Leftarrow \{\text{polynomials of degree} \leq n\}$, $p \neq p_*$.

We will say a few things about $p^*$ while doing this.

<u>Task 1</u>

<u>Theorem</u> (Weierstrass Approximation Theorem)

Let $f \in C[0,1]$ be a continuous function, and $\varepsilon > 0$. Then there is a polynomial $p$ with $\|f - p\| \leq \varepsilon$

<u>Proof</u> (Bernstein)

Let $n$ be a large integer (we'll choose $n \geq n_0(\varepsilon, f)$ ← depending on $\varepsilon, f$

Let $t \in [0,1]$. Let $X_1, \dots, X_n$ be independent Bernoulli random variables with mean $t$. Then $\mathbb{P}(X_i = 1) = t$, $\mathbb{P}(X_i = 0) = 1 - t$.

Let $Y_t = \frac{1}{n}\sum_{i=1}^{n} X_i$. We claim that $\mathbb{E}(f(Y_t))$ is a polynomial in $t$, and that it closely approximates $f(t)$. Idea: $Y_t$ is highly concentrated about $t$, so with high probability, $f(Y_t) \approx f(t)$.

By IA Probability theory, $\mathbb{P}(X_1 + \dots + X_n = k) = \binom{n}{k} t^k (1-t)^{n-k}$

Thus $\mathbb{E}(f(Y_t)) = \sum_{k=0}^{n} \mathbb{P}(Y_t = k/n) f(k/n)$

$= \sum_{k=0}^{n} \binom{n}{k} f(k/n) t^k (1-t)^{n-k}$ is a polynomial in $t$ of degree $n$.

Now we need to show that $P(t) = \mathbb{E}(f(Y_t))$ approximates $f(t)$ closely for large $n$. Let $\varepsilon > 0$. Choose $\delta > 0$ such that $|x - y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \varepsilon/2$ (possible since $f$ is uniformly continuous). Now $|p(t) - f(t)| = |\mathbb{E}(f(Y_t) - f(t))|$

$\leq \mathbb{E}(|f(Y_t) - f(t)|) \leq \varepsilon/2 + 2\|f\|\mathbb{P}$

30/01/13

# Topics in Analysis 6

$\mathbb{E}(|f(Y_t) - f(t)|) \le \varepsilon/2 + 2\|f\| \mathbb{P}(|Y_t - t| \ge \delta)$

Explanation: If $|Y_t - t| \le \delta$ then $|f(Y_t) - f(t)| \le \varepsilon/2$, otherwise, the trivial bound $|f(Y_t) - f(t)| \le 2\|f\|$.

To complete the proof, we need only show that if $n$ (dependent on $\varepsilon, f, \delta$) is large enough, then $\mathbb{P}(|Y_t - t| \ge \delta) \le \varepsilon/4\|f\|$.

We will actually show that $\mathbb{P}(|Y_t - t| \ge \delta) \le \frac{1}{\delta^2 n}$ which is definitely sufficient.

This is proved using Chebyshev's Inequality (IA Probability).

Recall that we define, for a random variable $X$ with $\mathbb{E}(X) = \mu$, $Var(X) = \underset{(*)}{\mathbb{E}((X-\mu)^2)}$. Also, if $X_1, \dots, X_n$ are independent, then $Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$ (by expanding $(*)$).

If $X_i$ is Bernoulli, mean $t$, then $Var(X_i) = t(1-t)$.

Then $Var(Y_t) = Var(\frac{1}{n}(X_1 + \dots + X_n)) = \frac{1}{n^2} Var(X_1 + \dots + X_n)$

$= \frac{1}{n} t(1-t) \le \frac{1}{n}$ (since $t(1-t) \le \frac{1}{4}$, $t \in [0,1]$)

Chebyshev's Inequality: $\mathbb{P}(|X - \mu| \ge \delta) \le \frac{Var(X)}{\delta^2}$

(Proof $Var(X) = \mathbb{E}(|X-\mu|^2) \ge \delta^2 \mathbb{P}(|X-\mu| \ge \delta)$ )

Putting this together gives the claimed bound. $\square$

## Task 2: Best Approximations

Writing $P_n = \{\text{polynomials of degree} \le n\}$.

### Theorem

Let $f: [0,1] \to \mathbb{R}$ be continuous. Then for each $n$, the minimum

$\inf_{p \in \mathcal{P}_n} \|f - p\|$ is attained. (N.B. not necessarily uniquely attained)
could be two such $p$

<u>Proof</u>

Consider $\Psi : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by $\Psi(a_0, a_1, \dots, a_n)$

$$= \sup_{t \in [0,1]} |f(t) - (a_0 + a_1 t + \dots + a_n t^n)| = \|f - P_{a_0, a_1, \dots, a_n}\|$$

where $P_{a_0, a_1, \dots, a_n} = a_0 + a_1 t + \dots + a_n t^n$. We claim that $\Psi$ is ~~continuous~~ continuous. Indeed $|P_{a_0,\dots,a_n}(t) - P_{a_0',\dots,a_n'}(t)| \leq |a_0 - a_0'| + \dots + |a_n - a_n'|$ for $t \in [0,1]$ by the triangle inequality.

Hence $|\Psi(a_0, \dots, a_n) - \Psi(a_0', \dots, a_n')| \leq |a_0 - a_0'| + \dots + |a_n - a_n'|$

$$\leq (n+1) \|(a_0, a_1, \dots, a_n), (a_0', \dots, a_n')\|_2$$

01/02/13 Topics in Analysis 7

<u>Weierstrass</u>

If $f \in C([0,1])$, $\varepsilon > 0$, then there is some polynomial $p$ with $\|f-p\| < \varepsilon$

<u>Recall</u>

$\mathcal{P}_n$ is the set of polynomials with degree $\leq n$ and real coefficients

<u>Theorem</u>

$\inf_{p \in \mathcal{P}_n} \|f-p\|$ is attained, for every $f \in C([0,1])$

<u>Proof</u>

Consider $\Psi : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by $\Psi(a_0, \dots, a_n) = \|f - P_{a_0,\dots,a_n}\|$ where $P_{a_0,\dots,a_n}(X) = a_0 + a_1 X + \dots + a_n X^n$. We showed that $\Psi$ is continuous. Now observe that there is some $c > 0$ such that $\|P_{a_0,\dots,a_n}\| \geq c$ uniformly for all $(a_0, \dots, a_n)$ on the unit ~~ball~~ sphere in $\mathbb{R}^{n+1}$. Indeed, $(a_0, \dots, a_n) \xrightarrow{\varphi} \|P_{a_0,\dots,a_n}\|$ is continuous and the unit ~~ball~~ sphere is compact. Furthermore, $0$ does not lie in the image of $\varphi$, since only the zero polynomial $P_{0,0,\dots,0}$ vanishes on $[0,1]$. Note, however, that $\|P_{\lambda a_0,\dots,\lambda a_n}\| = \lambda \|P_{a_0,\dots,a_n}\|$ and hence $\|P_{a_0,\dots,a_n}\| \to \infty$ as $(a_0,\dots,a_n) \to \infty$. It follows that there is some $R$ such that $\overset{①}{\|f - P_{a_0,\dots,a_n}\|} > \overset{②}{\inf_{p \in \mathcal{P}_n} \|f - p\|}$ whenever $\|(a_0,\dots,a_n)\| \geq R$. Indeed ① $\to \infty$ as $R \to \infty$, and ② is finite. Hence, $\inf_{p \in \mathcal{P}_n} \|f-p\| = \inf_{\|(a_0,\dots,a_n)\| \leq R} \|f - P_{a_0,\dots,a_n}\|$ which is the infimum of a continuous function on a compact set and is thus attained.

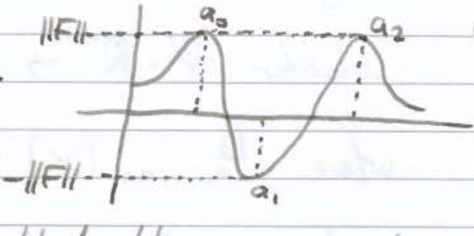
## Finding the Closest Polynomial to $f$ (best approximant)

How do we recognise that $p$ is (the) best approximant to $f$, with $p \in \mathbb{P}_n$

Certainly, $F := f - p$ has the property that $\|F + q\| \geq \|F\|$ whenever $q \in \mathbb{P}_n$. Such a function $F$ shall be referred to as <u>$n$-unimprovable</u>.

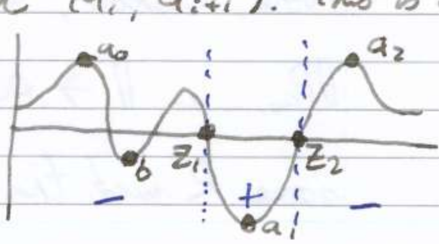
<u>Theorem</u> (Chebyshev 'Equal Ripple' Theorem)

A continuous function $F : [0,1] \to \mathbb{R}$ is <u>$n$-unimprovable</u> $\Longleftrightarrow$ there exist $n+2$ nodes $0 \leq a_0 < a_1 < \dots < a_{n+1} \leq 1$ such that $F(a_i) = \pm\|F\|$ with alternating signs.




<u>Proof</u>

The easier direction is 'if', ($\Leftarrow$). Suppose that there are nodes $a_0, \dots, a_{n+1}$. Suppose that $q \in \mathbb{P}_n$ is such that $\|F + q\| < \|F\|$. Suppose that $F(a_i) = (-1)^i \|F\|$ (the other case is identical). Then $q(a_0) < 0$ (otherwise $(F+q)(a_0) > \|F\|$), and $q(a_1) > 0$, $q(a_2) < 0$, and so on. So by the Intermediate Value Theorem, $q$ has at least $n+1$ roots, one in each interval $(a_i, a_{i+1})$. This is a contradiction as $q \in \mathbb{P}_n$ ※




* The 'only if', ($\Rightarrow$) direction is more tricky.

Let $F$ be an $n$-unimprovable function. Define $a_0$ to be the first point where $|F|$ first attains its maximum. WLOG, $F(a_0) = \|F\|$. Let $a_1$ be the first point after $a_0$ with $F(a_1) = -\|F\|$, then define $a_2, a_3$, and so on. We get points $a_0, \dots, a_m$. We must prove that $m \geq n+1$.

* NON-EXAMINABLE *

*NON-EXAMINABLE*

Suppose instead that $m \leq n$. We will show that $\exists q \in \mathcal{P}_n$ such that $\|F+q\| < \|F\|$, so $F$ wasn't $n$-unimprovable after all.

Let $z_i$ be the last point in $(a_{i-1}, a_i)$ where $F(z_i) = 0$.

Take $q(x) = \pm c(X - z_1)\dots(X - z_m)$ for some very small $c$.

Note that $q \in \mathcal{P}_n$. Choose the sign $\pm$ to match the diagram.

We claim that for sufficiently small $c$, $\|F+q\| < \|F\|$.

Certainly everything is fine at the extrema $a_i$, since the signs were chosen to make this so. Things could go wrong at points such as $b$, but we must have $F(b) \geq F(a_1) + \underset{(\dagger)}{\delta}$, $\delta > 0$, otherwise we would have chosen $b$ instead of $a_1$. Choosing $c$ small enough, we have $(F+q)(b) > F(a_1)$ $\square$

$(\dagger)$ N.B. On $[a_0, z_1]$, $F$ attains its minimum, which is <u>not</u> $F(a_1)$.

*NON-EXAMINABLE*

<u>Theorem</u>

Suppose that $f \in C[0,1]$. Then there is a unique best approximation to $f$ in $\mathcal{P}_n$.

<u>Proof</u>

Suppose that $f - p$ is a best approximation to $f$. Then $F$ is $n$-unimprovable. Thus $\|F+q\| \geq \|F\|$. We claim that this is a strict inequality if $q \neq 0$. Suppose that $\|F+q\| = \|F\|$.

By the triangle inequality, $\|F + q/2\| \leq \frac{1}{2}\|F\| + \frac{1}{2}\|F+q\| = \|F\|$.

So $F + q/2$ is also $n$-unimprovable.

By the equal ripple Criterion (difficult direction) there are nodes $a_0, \dots, a_{n+1}$ with the claimed properties.

Then $|(F+\frac{q}{2})(a_i)| = \|F\|$ for each $i$.

But $|(F+\frac{q}{2})(a_i)| \leq \frac{1}{2}|F(a_i)| + \frac{1}{2}\underbrace{|(F+q)(a_i)|}_{\leq \|F\|}$ (triangle inequality)

Equality must occur, so $(F+q/2)(a_i) = F(a_i) = (F+q)(a_i)$.

Hence $q(a_0) = \dots = q(a_{n+1}) = 0$. But $q \in P_n$, so $q \equiv 0$. $\square$

04/02/13

# Topics in Analysis ⑧

## Polynomial Interpolation

### Lagrange Interpolation

Given $n+1$ points $a_0, \dots, a_n \in \mathbb{R}$, and some function $f : \mathbb{R} \to \mathbb{R}$. Then we may construct explicitly a polynomial $p$ of degree $\leq n$ with $f(a_i) = p(a_i)$, $0 \leq i \leq n$. Indeed, consider $L_i(x) = \prod_{j \neq i} \frac{x - a_j}{a_i - a_j}$

Note that $L_i(a_j) = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$. Take $p(x) = \sum_{i=0}^{n} f(a_i) L_i(x)$

It is pretty obvious that $p$ has the stated properties.

### Basic Question

Is the polynomial $p$ constructed in this way a good approximation to $f$ in the uniform norm?

### Answer 1

No, not for any choice of the $a_i$, for particularly badly chosen $f$

### Answer 2

If $f$ is "nice" then this ~~choice~~ $p$ may be a good approximation to $f$, but some choices of the nodes $a_i$ are better than others. In particular equally spaced nodes are not great.

### Theorem

Suppose $a_0, \dots, a_n$ are distinct points in $[-1, 1]$ say. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable on some interval containing $[-1, 1]$. Let $p$ be the Lagrange Interpolant to $f$ of degree $\leq n$; that is to say the unique $p$ with degree $\leq n$ and $f(a_i) = p(a_i)$, $0 \leq i \leq n$.

Then for each $x \in (-1, 1)$, there is a $\xi \in (-1,1)$ such that
$f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-a_0)..(x-a_n)$
In particular $\|f-p\| \leq \frac{1}{(n+1)!}\|f^{(n+1)}\| \|(x-a_0)...(x-a_n)\|_{L^\infty(-1,1)}$

Proof

Define $r(t) = f(t) - p(t)$, and $w(t) = (t-a_0)...(t-a_n)$,
$y(t) = r(t) - \frac{w(t)}{w(x)} r(x)$. Note that $y(x) = 0$ (indeed, $w(a_i) = 0$
and $r(a_i) = f(a_i) - p(a_i) = 0$). Suppose that $x \notin \{a_0, ..., a_n\}$
(otherwise the result is trivial). Recall Rolle's Theorem:
If $f$ is differentiable and $f(a) = f(b) = 0$, $\exists c \in (a,b)$ such that
$f'(c) = 0$. We have a function $y$, with $n+2$ distinct zeroes.
Applying Rolle $n+1$ times, we get a $\xi$ where $y^{(n+1)}(\xi) = 0$.
Now $w^{(n+1)}(\xi) = (n+1)!$, since $w$ is a monic polynomial of degree $n+1$
Also, $r^{(n+1)}(\xi) = f^{(n+1)}(\xi)$, since $r = f - p$, and $p$ is a polynomial of degree $\leq n$. Hence $0 = y^{(n+1)}(\xi) = f^{(n+1)}(\xi) - \frac{(n+1)!\, r(x)}{w(x)}$
This is what was claimed. $\square$

Remark

When $n = 0$, this is the Mean Value Theorem.

Problem

How should we choose the nodes $a_0, ..., a_n$ to minimise $\max_{x \in [-1,1]} |(x-a_0)..(x-a_n)|$
and hence, in view of the Theorem, give a good approximation to $f$.

04/01/13 Topics in Analysis ⑧

<u>Chebyshev Polynomials</u>

<u>Definition</u>

$T_n$ is the unique polynomial (of degree $n$) such that $T_n(\cos\theta) = \cos n\theta$ for all $\theta$. We can calculate the first few by hand:

$\cos 2\theta = 2\cos^2\theta - 1$, $T_2(x) = 2x^2 - 1$

$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$, $T_3(x) = 4x^3 - 3x$

① $\cos((n+1)\theta) = \cos n\theta \cos\theta - \sin n\theta \sin\theta$

② $\cos((n+2)\theta) = \cos n\theta \cos 2\theta - \sin n\theta \sin 2\theta = \cos n\theta \cos 2\theta - 2\sin n\theta \sin\theta \cos\theta$

② - ① $\times 2\cos\theta \Rightarrow \cos((n+2)\theta) - 2\cos((n+1)\theta)\cos\theta = \cos n\theta \cos 2\theta - 2\cos^2\theta \cos n\theta$

$= -\cos n\theta$

Hence, if we know that $T_n$, $T_{n+1}$ exist then so does $T_{n+2}$, and $T_{n+2}(x) = 2xT_{n+1}(x) - T_n(x)$. So the Chebyshev polynomials are easy to compute and $T_n(x) = 2^{n-1}x^n + \ldots$

Note that for $x \in [-1,1]$, $|T_n(x)| \leq 1$, and the maximum is attained when $x = \cos\theta$, $\theta = \frac{k\pi}{n}$, $k = 0, 1, \ldots, n$. Write $a_k = \cos\frac{k\pi}{n}$ ~~(these are called the Chebyshev nodes)~~. Then $T_n(a_k) = (-1$

This means that $T_n$ is $(n-1)$-unimprovable, which means that $\|T_n + p\|_{L^\infty([-1,1])} \geq \|T_n\|_{L^\infty([-1,1])}$ whenever $p$ is a polynomial of degree $\leq n-1$.

Equivalently, $\frac{1}{2^{n-1}}T_n(x)$ has the smallest supremum norm on $[-1,1]$, amongst all monic polynomials of degree $\leq n$. Changing $n$ to $n+1$, we see that $\sup_{x \in [-1,1]} |(x-a_0)\ldots(x-a_n)| \geq \frac{1}{2^n}$ with equality when $b_0, \dots, b_n$ are the roots of the Chebyshev polynomial $T_{n+1}(x)$.

These are the Chebyshev nodes. These are given by $b_k = \cos\left[(k+\frac{1}{2})\pi / n+1\right]$

These give a good set of points for polynomial interpolation.

Runge's Phenomenon

With equally spaced interpolation points $a_0, \dots, a_n$, it is possible to have $\|f - p_n\| \to \infty$, where $p_n$ is the degree $\leq n$ interpolant to $f$ at the points $a_0, \dots, a_n$, even for "nice" $f$. Runge's example is $f(x) = \frac{1}{1+25x^2}$ on $[-1, 1]$. For $x = 0.9$ (say), $|f(x) - p_n(x)| \to \infty$.

Interpolating at the Chebyshev nodes is generally, but not always, more successful.

06/02/13

Topics in Analysis 9

## Quadrature and Legendre Polynomials

Problem

Given a continuous function $f: [-1, 1] \to \mathbb{R}$, approximate the integral $\int_{-1}^{1} f(t)\,dt$ by a discrete sum $\sum_{i=1}^{n} \lambda_i f(a_i)$ where the $\lambda_i \in \mathbb{R}$ are some weights and $-1 \leq a_1 < a_2 < \dots < a_n \leq 1$, are nodes.

We will show that by choosing $a_i$, $\lambda_i$ judiciously, this can be done.

It turns out that choosing $\lambda_i = \frac{1}{n}$, $a_i$ equally spaced is not a good choice.

Lemma

Given nodes $a_1, \dots, a_n$, there is a (unique) choice of weights $\lambda_i$ such that the "quadrature formula" $\int_{-1}^{1} f(t)\,dt = \sum_{i=1}^{n} \lambda_i f(a_i)$ whenever $f \in \mathcal{P}_{<n}$, the space of polynomials of degree $< n$.

Proof (Linear Algebra)

The map which sends $f$ to $\int_{-1}^{1} f(t)\,dt$ is a linear functional on the space $V = \mathcal{P}_{<n}$, that is to say it is in $V^*$. It suffices to show that the evaluation maps $f \mapsto f(a_i)$ give a basis for $V^*$. Since $\dim V^* = \dim V = n$, it suffices to show that these are linearly independent ie. $\sum_{i=1}^{n} \lambda_i f(a_i)$ for all $f \in V = \mathcal{P}_{<n}$

$\Rightarrow \lambda_1 = \dots = \lambda_n = 0$

Proof 1

Substitute $f(x) = x^j$, $j = 0, \dots, n-1$. We get

$$\begin{pmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_n \\ a_1^2 & \cdots & a_n^2 \\ \vdots & & \vdots \\ a_1^{n-1} & \cdots & a_n^{n-1} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = 0.$$ But the determinant of the matrix

is $\prod_{i \neq j} (a_i - a_j) \neq 0$ (Van der Monde determinant).

Hence $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$

Proof 2

Construct $f$ such that $f(a_i) = 1$ and $f(a_j) = 0$ for $j \neq i$, by Lagrange interpolation i.e. $f(x) = \prod_{j \neq i} \frac{x - a_j}{a_i - a_j} = L_i(x)$

The second proof gives a formula for the weights $\lambda_i$. Indeed, substituting $f(t) = L_i(t)$ into the quadrature formula gives $\lambda_i = \int_{-1}^{1} L_i(t)\,dt$. Thus, for any choice of the nodes $a_i$, we have a "reasonably sensible" choice of the weights $a_i$. It turns out that there is a particularly good choice of the $a_i$ coming from the Legendre polynomials.

Legendre Polynomials

Theorem

There are polynomials $p_n$, unique up to scalar multiples such that $\deg p_n = n$ and $\int_{-1}^{1} p_m(t)\,p_n(t)\,dt = 0$ if $m \neq n$.

If the scalar multiple is chosen so that $p_n(1) = 1$, these are called the Legendre Polynomials.

Proof

Apply the Gram-Schmidt process. Suppose that $p_0, p_1, \ldots, p_{n-1}$ have already been constructed. We can take $p_0(x) = 1$, $p_1(x) = x$

06/02/13

# Topics in Analysis ⑨

To make $p_n$, take a polynomial $F$ of degree $n$ (say $F(x) = x^n$) and define $p_n(x) = \sum_{i=0}^{n-1} \frac{\langle F, p_i \rangle}{\langle p_i, p_i \rangle} p_i$, where $\langle f, g \rangle$ denotes the inner product $\int_{-1}^{1} f(t)g(t)\,dt$.

$$\text{Then } \langle p_n, p_m \rangle = \langle F, p_m \rangle - \sum_{i=0}^{n-1} \frac{\langle F, p_i \rangle}{\langle p_i, p_i \rangle} \underbrace{\langle p_i, p_m \rangle}_{\text{non-zero only when } i=m}$$

$$= \langle F, p_m \rangle - \langle F, p_m \rangle = 0.$$

Note that $p_n$ has degree $n$. Why are the $p_i$ unique?

The polynomials $p_0, \ldots, p_{n-1}$ are linearly independent (either using the fact that orthogonal vectors are always linearly independent, or alternatively supposing that $\lambda_0 p_0 + \ldots + \lambda_{n-1} p_{n-1} = 0$, and considering the coefficients of $x^{n-1}$ to get $\lambda_{n-1} = 0$, inducting downwards).

Thus we <u>must</u> have $p_n = F - \sum_{i=0}^{n-1} \lambda_i p_i$. Now, taking inner products with $p_m$ gives $0 = \langle F, p_m \rangle - \lambda_m \langle p_m, p_m \rangle$

So indeed $\lambda_i = \frac{\langle F, p_i \rangle}{\langle p_i, p_i \rangle}$

<u>Lemma</u>

$p_n(x)$ has distinct roots $a_1 < \ldots < a_n$ in $[-1, 1]$.

Suppose not. Let $a_1, \ldots, a_k$ be all of the crossing roots of $p_n$. Suppose $k < n$, and consider $q(x) = (x - a_1)\ldots(x - a_k)$

Then $q p_n$ is always positive or always negative on $[-1, 1]$, and in particular $\int_{-1}^{1} q p_n(t)\,dt \neq 0$. But $p_n$ is orthogonal to all polynomials (such as $q$) of degree $< n$. ✱

Thus $p_n$ has $n$ distinct roots $a_1, \ldots, a_n$; the <u>Legendre nodes</u> (no explicit form)

Consider the quadrature formula attempt $\int_{-1}^{1} f(t)\,dt \approx \sum_{i=1}^{n} \lambda_i f(a_i)$ with these nodes, and the $\lambda_i$ chosen so that it is an exact formula for $f \in \mathcal{P}_{<n}$.

Claim

The quadrature "formula" is in fact exact for all $f \in \mathcal{P}_{<2n}$ i.e. for all polynomials of degree $2n-1$.

Proof

Given a polynomial $f \in \mathcal{P}_{<2n}$, we can write $f(x) = p_n(x)Q(x) + r(x)$ with $\deg Q \leq n-1$, $\deg r \leq n-1$. Then we have

$\int_{-1}^{1} f(t)\,dt = \int_{-1}^{1} r(t)\,dt$ (since $\langle p_n, Q\rangle = 0$, orthogonality)

Also $\sum_{i=1}^{n} \lambda_i f(a_i) = \sum_{i=1}^{n} \lambda_i \left(p_n(a_i)Q(a_i) + r(a_i)\right) = \sum_{i=1}^{n} \lambda_i r(a_i)$ (since $p_n(a_i) = 0$). But $\int_{-1}^{1} r(t)\,dt = \sum_{i=1}^{n} \lambda_i r(a_i)$ since $\deg r \leq n-1$. $\square$

Lemma

If the $\lambda_i$ are the weights attached to the Legendre nodes $a_1, \dots, a_n$ then $\sum_{i=1}^{n} \lambda_i = 2$ and $\lambda_i \geq 0$ for all $i$ (a crucial point).

Proof

For the first part, substitute $f \equiv 1$ into the quadrature formula (works for arbitrary nodes). For the second statement, $f(x) = \prod_{i \neq j} (x - a_j)^2$ has degree $2n-2$. Thus $0 < \int_{-1}^{1} f(t)\,dt = \sum_{i=1}^{n} \lambda_i f(a_i) = \prod_{j \neq i} (a_i - a_j)^2 \lambda_i$

Hence $\lambda_i > 0$. $\square$

08/02/13

# Topics in Analysis (10)

Legendre nodes $a_i$, $\quad -1 \leq a_1 < a_2 < \ldots < a_n \leq 1$

$$\int_{-1}^{1} f(t)\,dt \sim \sum_{i=1}^{n} \lambda_i f(a_i) \qquad \text{exact for } \deg f < n$$

When the $a_i$ are the Legendre nodes, exact for $\deg f < 2n$

In this case, $\sum_{i=1}^{n} \lambda_i = 2$, $\lambda_i > 0$.

__Proposition__

Let $f : [-1,1] \to \mathbb{R}$ be a continuous function. Then, as $n \to \infty$, $\sum_{i=1}^{n} \lambda_i f(a_i) \to \int_{-1}^{1} f(t)\,dt$

__Proof:__

This is obvious when $f$ is a polynomial, since the formula is exact when $n > \deg f$. In the general case, let $\varepsilon > 0$. By the Weierstrass Approximation Theorem, there is a polynomial $\tilde{f}$ such that $\|f - \tilde{f}\| < \varepsilon/10$. We have

$$\left| \int_{-1}^{1} f(t)\,dt - \int_{-1}^{1} \tilde{f}(t)\,dt \right| < \varepsilon/5 \quad (i)$$

$$\sum_{i=1}^{n} \lambda_i \tilde{f}(a_i) = \int_{-1}^{1} \tilde{f}(t)\,dt \quad \text{for sufficiently large } n \quad (ii)$$

$$\left| \sum_{i=1}^{n} \lambda_i f(a_i) - \sum_{i=1}^{n} \lambda_i \tilde{f}(a_i) \right| \leq \sum_{i=1}^{n} |\lambda_i| \, |f(a_i) - \tilde{f}(a_i)| \quad (iii)$$

$$< \varepsilon/10 \sum_{i=1}^{n} |\lambda_i| = \varepsilon/5 \quad \text{since } \lambda_i > 0, \ \sum \lambda_i = 2$$

Combining i), ii), iii), and using the Triangle Inequality gives

$$\left| \sum_{i=1}^{n} \lambda_i f(a_i) - \int_{-1}^{1} f(t)\,dt \right| < \frac{2\varepsilon}{5} < \varepsilon \quad \text{for sufficiently large } n$$

$\square$

__Remark__

The positivity of the $\lambda_i$ was crucial. When the $a_i$ are equally spaced, some of the $\lambda_i$ are negative, at least for $n \geq 8$. Indeed, it turns out that there are continuous $f$ for which the equally spaced quadrature formula diverges.

## Remark

On Example Sheet 3:

i) Compute a few Legendre Polynomials

ii) Recurrence relation linking $p_{n+2}, p_{n+1}, p_n$

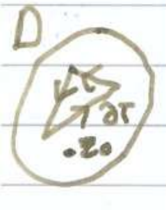
iii) Rodrigues Formula: $p_n(x) = \frac{1}{2^n n!}\left(\frac{d}{dx}\right)^n (1-x^2)^n$
(Hint: prove $\int_{-1}^{1} p_n(x)\, p_m(x)\, dx = 0$ by repeated integration by parts, then check $p_n(1) = 1$ )

## Approximation by complex polynomials

Basic Question: Suppose that $K \subseteq \mathbb{C}$ is compact, and $f: K \to \mathbb{C}$ continuous. Can $f$ be uniformly approximated by polynomials on $K$? That is, is it ~~possible~~ true that $\forall \varepsilon > 0$, there is a polynomial $p: \mathbb{C} \to \mathbb{C}$ such that $\sup_{z \in K} |p(z) - f(z)| \leq \varepsilon$
(Complex version of Weierstrass approximation)

There are two basic obstructions to such a result being true.

1. The uniform limit of holomorphic (complex differentiable functions) is holomorphic. More precisely, if $D \subseteq \mathbb{C}$ is an open ball, and if $f_n \to f$ uniformly on $D$, $f_n$ holomorphic, then $f$ is holomorphic.
Recall the Proof:




Morera's Theorem says that a continuous function $g: D \to \mathbb{C}$ is holomorphic $\iff \int_{\partial T} g(z)\, dz = 0$ for every triangle $T \subseteq D$.
One direction is via Cauchy's Theorem.

Conversely, define $G(z) = \int_{[z_0 \to z]} g(w)\, dw$. Then $G(z)$ is differentiable with derivative $g(z)$. Indeed,

$$\frac{G(z+h) - G(z)}{h} = \frac{1}{h}\int_{[z \to z+h]} g(w)\, dw \approx \frac{1}{h}\int_{[z \to z+h]} g(z)\, dw = g(z)$$

But holomorphic functions are infinitely differentiable, and so $g(z)$ is holomorphic. (Prove that something is differentiable by integrating it !!)

08/02/16

# Topics in Analysis ⑩

Returning to our discussion of obstruction 1, suppose that $\partial T$ is the boundary of a triangle $T$. Then

$$\left| \int_{\partial T} f_n(z)\,dz - \int_{\partial T} f(z)\,dz \right| \le \text{length}(\partial T)\, \|f_n - f\|_{L^\infty(D)} \to 0$$

since $f_n \to f$ uniformly. But $\int_{\partial T} f_n(z)\,dz = 0\ \forall n$, so $\int_{\partial T} f(z)\,dz = 0$. Hence $f$ is holomorphic. Note that $f$ was certainly continuous as the uniform limit of continuous functions.

2 "Topological Obstructions". Let $K = \{z \in \mathbb{C} \mid 1 \le |z| \le 2\}$. $f(z) = \frac{1}{z}$. $f(z)$ is holomorphic on $K$. However, $f$ is not the uniform limit of polynomials.

Indeed, let $\gamma$ be the contour $\gamma(t) = e^{2\pi i t}$, $t \in [0,1]$. If $p$ is a polynomial, $\int_\gamma p(z)\,dz = 0$ (Cauchy's Theorem).

However, $\int_\gamma \frac{1}{z}\,dz = \int_0^1 \frac{\gamma'(t)}{\gamma(t)}\,dt = 2\pi i$.

But $\int_\gamma \left(p(z) - \frac{1}{z}\right) dz \le \text{length}(\gamma)\, \|p - \frac{1}{z}\|_{L^\infty(K)}$ and so $\|p - \frac{1}{z}\|_{L^\infty(K)} \ge 1$.

More generally, there can be no "holes" in $f$.

In a sense, these are the only ways in which a function $f : K \to \mathbb{C}$ can fail to be uniformly approximated by polynomia[ls]. This is called Runge's Theorem.

<u>Theorem</u>

$K \subseteq \Omega$

Suppose that $K$ is compact and that $\Omega$ is a bounded domain, open in $\mathbb{C}$, such that $\mathbb{C} \setminus \Omega$ is path connected. Then, let $f : \Omega \to \mathbb{C}$ be holomorphic. Then $f$ can be uniformly approximate[d]

Path connected means that any two points in $\mathbb{C} \setminus \Omega$ are joined by a path entirely in $\mathbb{C} \setminus \Omega$.

Broad Strategy

1. $f$ can be uniformly approximated by rational functions, all of whose poles lie outside of $K$.

   Then we will reduce to the case $f(z) = \frac{1}{a-z}$, $a \notin K$

2. Handle $f(z) = \frac{1}{a-z}$, $a \notin K$.

## Topics in Analysis (11)

### Runge's Theorem

**Theorem**

Suppose that $K \subseteq \mathbb{C}$ is compact, $\Omega \subseteq \mathbb{C}$ open, and $K \subseteq \Omega$. Suppose that $f : \Omega \to \mathbb{C}$ is analytic, and $\mathbb{C} \setminus K$ path connected. Then $f$ may be uniformly approximated by polynomials.

"Can be uniformly approximated by polynomials on $K$" $\Leftrightarrow$ "Is UAP"

We observe that if $f, g : K \to \mathbb{C}$ are UAP then so are $f+g$, $fg$, and $\lambda f$, for any scalar $\lambda$ (Proof on the Example Sheet. The key point is that $f$, $g$ are bounded on $K$).

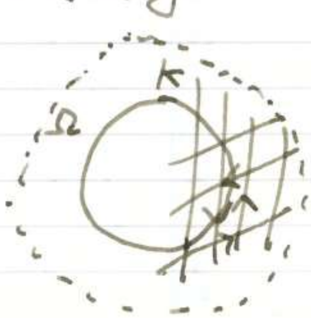
**Proof of Step 1**

Recall Cauchy's Integral Formula. If $f$ is holomorphic on a square domain $D$, and if $\gamma$ is a small square contour then $f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{z-w}\,d$ whenever $z \in \gamma$.




(Brief reminder of proof : Shrink the contour to a very small circular contour about $z$, $C_\varepsilon(z)$. If $\varepsilon$ is very small then $f(w) \approx f(z)$. The integral does not change.)

Since $K$ is compact, and $\mathbb{C} \setminus \Omega$ is closed, there is some $\delta > 0$ such that $|x - y| \geq \delta$ whenever $x \in K$, $y \notin \Omega$ (Sheet 1, Question 2). Place on the complex plane a square grid, sidelength $\delta/100$, centred on $0$. Apply Cauchy's Integral Formula around every subsquare in the grid which touches $K$.




Write $c_1, c_2, \dots, c_m$ for all of the square contours. Provided that $z$ does not lie on any $c_j$, we have $f(z) = \frac{1}{2\pi i}\sum_{j=1}^{m}\int_{c_j}\frac{f(w)}{w-z}\,dz$.

Why? If $z$ lies inside $c_j$, apply Cauchy's Integral Formula. If $z$ does not lie in $c_j$, then $\int_{c_j}\frac{f(w)}{w-z}\,dz = 0$, since $\frac{f(w)}{w-z}$ is holomorphic on ~~a~~ an open domain containing $c_j$.

But every $z \in K$ does lie in or on the boundary of one of the $c_i$.

When the sum over $j$ is taken, a lot of cancellation occurs. Any edge ~~completely in~~ meeting $K$ will be cancelled out.

Thus we obtain $f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}\,dw$, where $\gamma$ is some closed contour lying between $K$ and $\mathbb{C}\setminus\Omega$, made up of horizontal or vertical segments.

Parametrise this path $\gamma$ as a continuous path $\varphi : [0,1] \to \mathbb{C}$. Then we have $f(z) = \frac{1}{2\pi i}\int_0^1 \frac{f(\varphi(t))}{\varphi(t)-z}\varphi'(t)\,dt$ (definition of a path integral). Approximate the integral by a sum at the points $t = \frac{j}{N}$, $j = 0, 1, \ldots, N-1$ for some large $N$.

Write this as $S_N(z) = \frac{1}{2\pi i}\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{f(\varphi(j/N))}{\varphi(j/N)-z}\varphi'(j/N)$. We will show in just a moment that $S_N(z) \to f(z)$ uniformly for $z \in K$. This completes the proof of Step 1, since the sum consists of rational functions with poles outside $K$. (The poles are at $\varphi(j/N)$, which lie on $\gamma$, and hence outside $K$).

* NON-EXAMINABLE

Why does $S_N(z) \to f(z)$ uniformly? Suppose that $F : [0,1] \to \mathbb{C}$ is continuous. Then $\left|\int_0^1 F(t)\,dt - \frac{1}{N}\sum_{j=1}^{N}F(j/N)\right| \le \sup_{|x-y|\le\frac{1}{N}}|F(x)-F(y)| \to 0$, as $F$ is uniformly continuous.

In our case, $F(t) = \frac{f(\varphi(t))}{\varphi(t)-z}\varphi'(t)$.

* NON EXAMINABLE

However, $\sup_{|x-y| \leq \frac{1}{N}} |F_z(x) - F_z(y)| \to 0$ uniformly in $z$, since the function $(z,t) \mapsto F_z(t)$ is uniformly continuous on $K \times [0,1]$, since $K$, $[0,1]$ are compact, and the map is continuous because $\varphi(t) \notin K$.

When $z$ lies on a square $C_j$. We got the formula
$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-z} dz$ when $z \in K$ does not lie on the boundary of a $C_j$. However, both sides are continuous functions of $z$ and so the formula also holds when $z$ lies on the boundary of one of the $C_j$. *

Step 2

We now want to show that if $a \notin K$, then $\frac{1}{a-z}$ is UAP on $K$.
If we can do this, then the proof of Runge is complete.
The idea is to consider the set $S$ of $a$ for which this is true.
We will show i) that $S$ is non-empty (easy). Then ii) if $a \in S$, and $b$ is "close" to $a$, then $b \in S$. We will conclude iii) that $S = \mathbb{C} \setminus K$.

Proof of i)

Take $a$ with $|a| > 2 \sup_{z \in K} |z|$. Then

$$\frac{1}{a-z} = \frac{1}{a}\left(\frac{1}{1-z/a}\right) = \frac{1}{a}\underbrace{\left(1 + \frac{z}{a} + \frac{z^2}{a^2} + \dots\right)}$$

Series converges uniformly on $z \in K$, since

$$\sum_{n \geq N} \left|\frac{z}{a}\right|^n \leq \sum_{n \geq N} \frac{1}{2^n} = \frac{1}{2^N}$$

13/02/11 Topics in Analysis (12)

Notes available at dpmms.cam.ac.uk/~bjg23

Runge's Theorem (Second part of the proof)

$K \subseteq \mathbb{C}$ compact, $\mathbb{C} \setminus K$ path connected. Show that $\frac{1}{\alpha - z}$ is UAP if $\alpha \notin K$.

Strategy

$\hat{S} = \{\alpha : \frac{1}{\alpha - z} \text{ is UAP}\}$

i) $\hat{S}$ non-empty (done last time)

ii) If $\alpha \in \hat{S}$, and $\beta$ is close to $\alpha$, then $\beta \in \hat{S}$.

iii) $\hat{S}$ is all of $\mathbb{C} \setminus K$.

ii) Note that $\frac{1}{\beta - z} = \frac{1}{(\alpha - z)\left(1 + \frac{\beta - \alpha}{\alpha - z}\right)}$. By the geometric series expansion, we have

$$\frac{1}{\beta - z} = \frac{1}{\alpha - z}\left(1 + \left(\frac{\alpha - \beta}{\alpha - z}\right) + \left(\frac{\alpha - \beta}{\alpha - z}\right)^2 + \dots\right)$$

If $|\alpha - \beta| < |\alpha - z|$, certainly if $|\alpha - \beta| < \operatorname{dist}(\alpha, K) = \inf_{z \in K} |\alpha - z|$ then the series converges uniformly for $z \in K$ (use the triangle inequality)

Write $S_N(z) = \frac{1}{\alpha - z}\left(1 + \left(\frac{\alpha - \beta}{\alpha - z}\right) + \dots + \left(\frac{\alpha - \beta}{\alpha - z}\right)^N\right)$. By earlier remarks, $S_N(z)$ is UAP, since, by assumption, $\frac{1}{\alpha - z}$ is UAP, and $S_N(z)$ can be built up using just sums, products, and scalar multiples.

Now $S_N(z) \to \frac{1}{\beta - z}$ uniformly for $z \in K$.

(indeed $\left|S_N(z) - \frac{1}{\beta - z}\right| \leq \frac{1}{|\alpha - z|} \sum_{n > N} \left|\frac{\alpha - \beta}{\alpha - z}\right|^n \leq \frac{1}{\operatorname{dist}(\alpha, K)} \sum_{n > N} r^n \to 0$ as $N \to \infty$

(where $r = \sup_{z \in K} \left(\frac{|\alpha - \beta|}{|\alpha - z|}\right) < 1$ )

Finally, note that a uniform limit of UAP functions is UAP.

(Proof: If $f_n \to f$ uniformly, choose $p_n$ such that

$\sup_{K \ni z} |p_n(z) - f_n(z)| < \frac{1}{n}$. Then $p_n \to f$ uniformly too.)

We have established the following precise version of ii) :

If $\alpha \in S$, $|\beta - \alpha| < \text{dist}(\alpha, K)$, then $\beta \in S$.

iii) Suppose that $\beta \in \mathbb{C} \setminus K$ is arbitrary.

Let $\alpha \in S$ (such an $\alpha$ exists by i) ). Consider a continuous path from $\alpha$ to $\beta$, that is to say, a continuous map $\varphi : [0,1] \to \mathbb{C} \setminus K$ with $\varphi(0) = \alpha$, $\varphi(1) = \beta$. The image $\varphi([0,1])$ is compact, since $[0,1]$ is compact, and hence closed.


Hence (Sheet 1, Q2), $\exists \delta > 0$ such that $\text{dist}(\varphi(t), K) \geq \delta$ uniformly for $t \in [0,1]$.

Since $[0,1]$ is compact, $\varphi$ is uniformly continuous, hence there is $\eta > 0$ such that $|x - y| \leq \eta \Rightarrow |\varphi(x) - \varphi(y)| < \delta$.

Take a finite sequence $0 = t_0 < t_1 < .. < t_m = 1$, with $t_{i+1} - t_i \leq \eta$ for each $i$. Then, for each $i$, $\text{dist}(\varphi(t_{i+1}), \varphi(t_i)) < \delta \leq \text{dist}(\varphi(t_i), K)$. Hence by ii), $\varphi(t_i) \in S \Rightarrow \varphi(t_{i+1}) \in S$. Hence, since $\varphi(t_0) = \alpha \in S$, we know $\varphi(t_m) = \beta \in S$ $\square$

<u>Theorem</u>

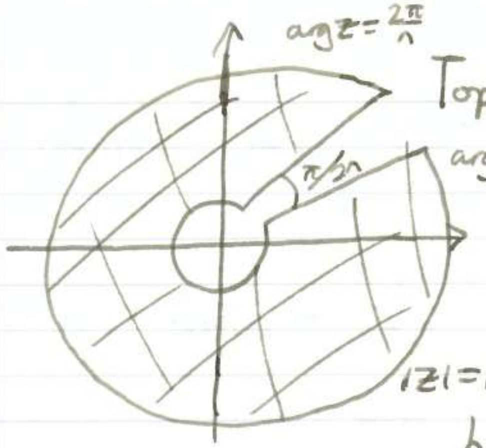
There exists a sequence of polynomials $p_n$ converging ~~uniformly~~ pointwise to a discontinuous function on $|z| \leq 1$.

<u>Proof</u>

Consider the following compact sets $K_n$.

13/02/13

# Topics in Analysis (12)




$K_n$

$= \{ z \in \mathbb{C} \mid |z| \leq 1,\ |z| \geq \frac{1}{n},\ \arg z \in [\frac{2\pi}{2n}, \frac{2\pi}{n}] \}$

On $K_n$, we can define a continuous branch of $\sqrt{z}$.

$f_n(re^{i\theta}) = r^{\frac{1}{2}} e^{i\theta/2}$ where $\frac{2\pi}{n} \leq \theta \leq 2\pi + \frac{2\pi}{2n}$

This is holomorphic on an open set $\Omega$, with $K_n \subseteq \Omega$, and $\mathbb{C} \setminus K_n$ is path-connected. By Runge, there is a polynomial $p_n$ such that $\sup_{z \in K_n} |p_n(z) - f_n(z)| \leq \frac{1}{n}$. Finally, consider the polynomia[l] $q_n(z) = z p_n(z)$.

Then $q_n(0) = 0$ for all $n$. If $z \neq 0$, then $z \in K_n$ for all $n$ sufficiently large. Hence $\lim_{n\to\infty} q_n(z) = \lim_{n\to\infty} z f_n(z)$
$= r^{3/2} e^{i\,3\theta/2}$ for $z = re^{i\theta}$, where $0 < \theta \leq 2\pi$.

This is discontinuous on the positive real line. □

## Chapter 3 : Approximation by Rationals

Contents

i) $e, \pi$ irrational

ii) Continued Fractions (hopefully including $e$)

iii) Transcendental Numbers (maybe proving that $e, \pi$ are transcendental) — NON-EXAMINABLE —

**Theorem**

$e$ is irrational.

**Proof**

$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \dots$. If $e = p/q$, consider $q!e \in$

But $q!e = q! + q! + \frac{q!}{2!} + \dots + \frac{q!}{q!} + \frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \dots$

Hence $\frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \dots \in \mathbb{N}$.

But $0 < \frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \dots < \frac{1}{q+1} + \frac{1}{(q+1)^2} + \dots = \frac{1}{q} < 1$

But there are no integers between 0 and 1. $\square$

<u>Theorem</u>

$\pi$ is irrational.

<u>Proof</u>

Suppose that $\pi = p/q$. Consider the polynomial $f_n(x) = \frac{q^n x^n (\pi - x)^n}{n!}$

<u>Sketch Proof</u> (to be expanded next time)

Consider $\int_0^\pi f_n(x) \sin x \, dx$. Integrate by parts repeatedly, until we get to $f_n^{(2n+1)} \equiv 0$. Conclude that

$\int_0^\pi f_n(x) \sin x \, dx \in \mathbb{Z}$.

But $0 < \int_0^\pi f_n(x) \sin x \, dx < \frac{q^n (\frac{\pi}{2})^{2n}}{n!} \pi \to 0$ as $n \to \infty$

$\#$

Theorem

$\pi$ is irrational.

Proof

Suppose that $\pi = p/q$, and for $n \in \mathbb{N}$, consider $f_n(x) = \frac{q^n x^n (\pi - x)^n}{n!}$

Observe that $f_n^{(m)}(0) = f_n^{(m)}(\pi) = 0$ for $m < n$, since $f_n$ vanishes to order $n$ at $0, \pi$). Also, $f_n^{(m)}(0) = f_n^{(m)}(\pi) = 0$ for $m > 2n$ (since $f_n$ is a polynomial).

Furthermore, $f_n^{(m)}(0), f_n^{(m)}(\pi) \in \mathbb{Z}$ for $m \geq n$, because by differentiating $x^r$ $n$ times we get $r(r-1)\dots(r-n+1)x^{r-n}$ or $0$. If $r \geq n$, then $n! \mid r(r-1)\dots(r-n+1)$ (since $\binom{n}{r} = \frac{r(r-1)\dots(r-n+1)}{n!}$)

and the denominators of $q^n$ coming from $\pi$ are cancelled by the factor of $q^n$.

Now look at $\int_0^\pi f_n(x) \sin x \, dx$. One may evaluate this using many integrations by parts (keep differentiating $f_n$ and integrating $\sin x$). This will terminate after $2n+1$ steps since $f^{(2n+1)}(x) = 0$

Along the way, we will be evaluating expressions like $\left[\pm f_n^{(m)}(x) \begin{Bmatrix} \cos x \\ \sin x \end{Bmatrix}\right]_0^\pi$. All of these are integers, since $f_n^{(m)}(0), f_n^{(m)}(\pi) \in \mathbb{Z}$.

Hence, $\int_0^\pi f_n(x) \sin x \, dx \in \mathbb{Z}$. However, $\int_0^\pi f_n(x) \sin x \, dx > 0$

Also, since $x(\pi - x) \leq \left(\frac{\pi}{2}\right)^2$, $\forall x \in [0, \pi]$, we have $f_n(x) \leq \frac{C^n}{n!}$ where $C = q\left(\frac{\pi}{2}\right)^2$, which tends to $0$ as $n \to \infty$

Thus, if $n$ is large, $\int_0^\pi f_n(x) \sin x \, dx < 1$ $\#$

## Continued Fractions

A <u>continued fraction</u> is an expression of the form

$a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots}}$ , $a_0, a_1, a_2, \ldots \in \mathbb{N}$.

i) Does this expression make sense / converge?

ii) Can we expand any real number like this?

iii) Are there some numbers whose continued fraction expansion is nice?

This is intimately related to the issue of approximating real numbers by rationals.

Notation: $[a_0, a_1, a_2, \ldots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots}}$

We will look at finite truncations $[a_0, \ldots, a_k]$.

**Example**

$[1,1,1,1] = 1 + \frac{1}{1+\frac{1}{1+1}} = \frac{5}{3}$ , $[1,1,1,1,1] = \frac{8}{5}$ , then $\frac{13}{8}, \frac{21}{13}$, etc.

This appears to be related to the sequence of Fibonacci numbers.

Thus if $\frac{p_k}{q_k} = [a_0, a_1, \ldots, a_k]$, then, when $a_i = 1 \ \forall i$, we appear to have $p_k = p_{k-1} + p_{k-2}$, $q_k = q_{k-1} + q_{k-2}$

**Proposition**

Suppose that $a_0, a_1, a_2, \ldots \in \mathbb{N}$. Write $\frac{p_k}{q_k} = [a_0, a_1, \ldots, a_k]$ in lowest terms. Then $p_k = a_k p_{k-1} + p_{k-2}$, $q_k = a_k q_{k-1} + q_{k-2}$

**Proof**

Observe that if $\frac{p'}{q'} = a_i + \frac{1}{(p/q)}$ (the continued fraction can be evaluated by k operations of this type), then

$\frac{p'}{q'} = a_i + \frac{q}{p} = \frac{a_i p + q}{p}$. Hence

$\begin{pmatrix} p' \\ q' \end{pmatrix} = \begin{pmatrix} a_i & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}$. Note that if $p, q$ have no common fact[or] neither do $p', q'$

(if $d|p'$ and $d|q'$ then $d | p' - a_i q' = p$, and $d|q$)

We start with $\frac{p}{q} = \frac{a_k}{1}$, already in lowest terms.

It follows that $\frac{p_k}{q_k} = [a_0, a_1, \dots, a_k]$ is given by

$$\begin{pmatrix}p_k\\q_k\end{pmatrix} = \underbrace{\begin{pmatrix}a_0&1\\1&0\end{pmatrix}\begin{pmatrix}a_1&1\\1&0\end{pmatrix}\cdots\begin{pmatrix}a_k&1\\1&0\end{pmatrix}}\begin{pmatrix}1\\0\end{pmatrix}$$

Write $M$ instead for $\begin{pmatrix}a_0&1\\1&0\end{pmatrix}\cdots\begin{pmatrix}a_{k-2}&1\\1&0\end{pmatrix}$

$$\begin{pmatrix}p_{k-2}\\q_{k-2}\end{pmatrix} = M\begin{pmatrix}1\\0\end{pmatrix}$$

$$\begin{pmatrix}p_{k-1}\\q_{k-1}\end{pmatrix} = M\begin{pmatrix}a_{k-1}&1\\1&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = M\begin{pmatrix}a_{k-1}\\1\end{pmatrix}$$

$$\begin{pmatrix}p_k\\q_k\end{pmatrix} = M\begin{pmatrix}a_{k-1}&1\\1&0\end{pmatrix}\begin{pmatrix}a_k&1\\1&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = M\begin{pmatrix}a_{k-1}a_k+1\\a_k\end{pmatrix}$$

Hence $\begin{pmatrix}p_k\\q_k\end{pmatrix} = a_k\begin{pmatrix}p_{k-1}\\q_{k-1}\end{pmatrix} + \begin{pmatrix}p_{k-2}\\q_{k-2}\end{pmatrix}$ □

Note that

$$\begin{pmatrix}p_k&p_{k+1}\\q_k&q_{k+1}\end{pmatrix} = \begin{pmatrix}a_0&1\\1&0\end{pmatrix}\cdots\begin{pmatrix}a_k&1\\1&0\end{pmatrix}\begin{pmatrix}1&a_{k+1}\\0&1\end{pmatrix}$$ (follows as before)

Taking determinants, $p_k q_{k+1} - p_{k+1} q_k = (-1)^{k+1}$

This implies that $\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}} = \frac{(-1)^{k+1}}{q_k q_{k+1}}$ (*)

Note that from the recurrence relation $q_k = a_k q_{k-1} + q_{k-2}$, we have $q_k \to \infty$ as $k \to \infty$ (in fact, $q_k$ grows at least as fast as Fibonacci, i.e. exponentially). Thus $\left|\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}}\right| \leq \frac{1}{k^2}$. Since $\sum \frac{1}{k^2}$ converges, $\frac{p_k}{q_k}$ is a Cauchy sequence, and hence converges. We have shown that if $a_0, a_1, a_2, \dots \in \mathbb{N}$, then $\lim_{n\to\infty}[a_0, \dots, a_n]$ exists. We denote this by

$[a_0, a_1, a_2, \dots]$ or $a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}}$

Fact

$[1, 1, 1, \dots] = 1 + \frac{1}{1 + \frac{1}{1 + \dots}} = \frac{1+\sqrt{5}}{2}$

Proof:

$\overbrace{[1,1,\ldots,1]}^{k \text{ terms}}{}_k = 1 + \frac{1}{[1,1,\ldots,1]_{k-1}}$. If $\alpha = \lim_{k\to\infty} [1,1,\ldots,1]_k$

then $\alpha = 1 + \frac{1}{\alpha}$. Solving the quadratic equation gives the result

since $\alpha = [1,1,\ldots]$. $\square$

Also, for example, $[1,2,2,\ldots] = \sqrt{2}$ and many similar expressions.

(*) immediately tells us that $\frac{p_0}{q_0} < \frac{p_1}{q_1}$, $\frac{p_1}{q_1} > \frac{p_2}{q_2}$, $\frac{p_2}{q_2} < \frac{p_3}{q_3}$ etc

Also, $\frac{p_k}{q_k} - \frac{p_{k+2}}{q_{k+2}} = \frac{(-1)^{k+1}}{q_k q_{k+1}} + \frac{(-1)^{k+2}}{q_{k+1} q_{k+2}} = \frac{(-1)}{q_{k+1}}\left(\frac{1}{q_k} - \frac{1}{q_{k+2}}\right)$

So $\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \ldots$ and $\frac{p_1}{q_1} > \frac{p_3}{q_3} > \ldots$

## Continued Fractions

Topics in Analysis (14)

$p_k = a_k p_{k-1} + p_{k-2} \qquad q_k = a_k q_{k-1} + q_{k-2}$

$\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}} = \frac{(-1)^k}{q_k q_{k+1}}$

## Representing a Real Number $(\alpha \in (0,\infty))$

Observe that for any $k$, one can write $\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots \frac{1}{a_k + \varepsilon_k}}}$ where $0 \leq \varepsilon_k < 1$. Indeed, to find the sequence of $a_i$, repeat the following two operations:

- Take integer parts e.g. $a_0 = \lfloor \alpha \rfloor$. Subtract this
- Reciprocate $(x \mapsto \frac{1}{x})$

Note that $\varepsilon_k = \frac{1}{a_{k+1} + \varepsilon_{k+1}}$, hence $\varepsilon_{k+1} = \{\frac{1}{\varepsilon_k}\}$ Fractional Part $= \frac{1}{\varepsilon_k} - \lfloor \frac{1}{\varepsilon_k} \rfloor$

The map $T: [0,1) \to [0,1)$, $x \mapsto \{\frac{1}{x}\}$ is called the Gauss Map.

Write $p_k/q_k = [a_0, \dots, a_k]$, $\frac{\tilde{p}_k}{\tilde{q}_k} = [a_0, \dots, a_k + \varepsilon_k] = \alpha$ ← not necessarily integers

We saw last time that

$\begin{pmatrix} p_k \\ q_k \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \dots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k \\ 1 \end{pmatrix}$

Similarly, $\begin{pmatrix} \tilde{p}_k \\ \tilde{q}_k \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \dots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k + \varepsilon_k \\ 1 \end{pmatrix}$

Hence $\begin{pmatrix} p_k & \tilde{p}_k \\ q_k & \tilde{q}_k \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \dots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k & a_k + \varepsilon_k \\ 1 & 1 \end{pmatrix}$

Taking determinants gives $|p_k \tilde{q}_k - q_k \tilde{p}_k| \leq 1$

Hence $|\alpha - \frac{p_k}{q_k}| = |\frac{\tilde{p}_k}{\tilde{q}_k} - \frac{p_k}{q_k}| \leq \frac{1}{q_k \tilde{q}_k} \to 0$

(Use the recurrence relations for $q_k, \tilde{q}_k$ to see that they $\to \infty$)

## Conclusion

Every positive real number has a continued fraction expansion, which is unique (exercise).

Further Properties of the Convergents

From the fact that $\frac{p_0}{q_0} < \frac{p_2}{q_2} < \dots < \alpha < \dots < \frac{p_3}{q_3} < \frac{p_1}{q_1}$ and the fact that $\left|\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}}\right| = \frac{1}{q_k q_{k+1}}$ we see that $\left|\alpha - \frac{p_k}{q_k}\right| \leq \frac{1}{q_k q_{k+1}}$. In particular, since $q_{k+1} \geq q_k$, we have $\left|\alpha - \frac{p_k}{q_k}\right| \leq \frac{1}{q_k^2}$ (so convergents tend to $\alpha$ rapidly)

Theorem (Best Rational Approximation)

Suppose that $q \leq q_k$. Then $\left|\alpha - \frac{p}{q}\right| \geq \inf\left(\left|\alpha - \frac{p_k}{q_k}\right|, \left|\alpha - \frac{p_{k-1}}{q_{k-1}}\right|\right)$ with equality only if $q = q_k$, $p = p_k$ or $q = q_{k-1}$, $p = p_{k-1}$

Proof

We have $\left|\alpha - \frac{p_k}{q_k}\right| \leq \frac{1}{q_k q_{k+1}}$. However, if $\frac{p}{q} \neq \frac{p_k}{q_k}$ (WLOG, $p/q$ is in lowest terms), then $\left|\frac{p}{q} - \frac{p_k}{q_k}\right| = \left|\frac{p q_k - q p_k}{q_k q}\right| \geq \frac{1}{q q_k} > \frac{1}{q_k q_{k+1}}$. Hence $p/q$, if it is closer to $\alpha$ than $\frac{p_k}{q_k}$, is on the other side of $\alpha$ from $\frac{p_k}{q_k}$.

~~Furthermore, if $\frac{p_{k-1}}{q_{k-1}}$~~ We could have $\frac{p_{k-1}}{q_{k-1}} = p/q$, in which case we are done. If not, $\left|\frac{p}{q} - \frac{p_{k-1}}{q_{k-1}}\right| = \left|\frac{p q_{k-1} - q p_{k-1}}{q q_{k-1}}\right| \geq \frac{1}{q q_{k-1}} \geq \frac{1}{q_{k-1} q_k}$

Hence $\left|\frac{p}{q} - \frac{p_{k-1}}{q_{k-1}}\right| \geq \left|\alpha - \frac{p_{k-1}}{q_{k-1}}\right|$, so $p/q$ is equal to $\alpha$ or is not between $\frac{p_{k-1}}{q_{k-1}}$ and $\alpha$. This covers all cases. $\square$

We have seen that convergents $p_k/q_k$ are good approximations to $\alpha$ in the sense that $\left|\alpha - p_k/q_k\right| \leq \frac{1}{q_k^2}$

Remarkably, the converse is true up to a factor of 2.

Theorem

Let $\alpha \in (0, \infty)$, $\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{2q^2}$, $p/q$ in lowest terms.

Then $p/q$ is one of the convergents to the continued fraction of $\alpha$.

Proof

Since the sequence of denominators $q_k$ is increasing, we choose $k$ such that $q_k \leq q < q_{k+1}$.

If $p/q = p_k/q_k$ we are done, so suppose not. Then

$$\left|p/q - p_k/q_k\right| \geq \frac{1}{q q_k} > \frac{1}{q_k q_{k+1}} \geqslant \left|\alpha - p_k/q_k\right|$$

Case 1

$p/q > p_k/q_k$. Then $p/q$ cannot lie between $p_k/q_k$ and $\alpha$.

Hence we may assume (the other case being similar) that $p_k/q_k < \alpha < p/q$. By the fact that $\frac{p_{k+1}}{q_{k+1}}$ is a better approximation to $\alpha$ than $p/q$, this implies that $p_k/q_k < \alpha < \frac{p_{k+1}}{q_{k+1}} < p/q$

Consider first of all the case $q \geq \frac{1}{2} q_{k+1}$. Then

$$\left|p/q - \alpha\right| \geq \left|\frac{p}{q} - \frac{p_{k+1}}{q_{k+1}}\right| \geq \frac{1}{q q_{k+1}} \geq \frac{1}{2q^2} \quad \#$$

Otherwise, suppose that $q < \frac{1}{2} q_{k+1}$. Then

$$\left|\frac{p}{q} - \alpha\right| \geq \left|\frac{p_k}{q_k} - \frac{p}{q}\right| - \left|\frac{p_k}{q_k} - \alpha\right| \geq \frac{1}{q q_k} - \frac{1}{q_k q_{k+1}}$$

$$= \frac{1}{q_k}\left(\frac{1}{q} - \frac{1}{q_{k+1}}\right) > \frac{1}{2 q_k q} \geq \frac{1}{2q^2}$$

**Two slightly tricky statements about continued fractions**

Let $\alpha$ be a real number, and let $[a_0, a_1, a_2, \dots]$ be its continued fraction expansion. Write $\frac{p_k}{q_k}$ for the convergents. To avoid annoyances, I'll assume that this expansion is infinite. This is so if and only if $\alpha$ is irrational. The less obvious direction is the *only if* direction, which is the statement that the continued fraction expansion of a rational number is finite. To see this, suppose that $\alpha = \frac{p}{q}$ is rational and has an infinite continued fraction expansion. Certainly, for each $k$, we have $\frac{p}{q} \neq \frac{p_k}{q_k}$, and so

$$|\alpha - \frac{p_k}{q_k}| = |\frac{p}{q} - \frac{p_k}{q_k}| \geqslant \frac{1}{qq_k}.$$

However we also know that $|\alpha - \frac{p_k}{q_k}| \leqslant \frac{1}{q_k^2}$. Hence $q_k \leqslant q$ for all $k$, a contradiction.

We proved in lectures that

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \alpha < \cdots < \frac{p_3}{q_3} < \frac{p_1}{q_1}.$$

Also, remember, we have

$$|\alpha - \frac{p_k}{q_k}| < \frac{1}{q_k q_{k+1}}$$

for all $k$.

We will make very frequent use of the following observation: if $\frac{a}{b}$ and $\frac{a'}{b'}$ are distinct fractions in lowest terms, then

$$|\frac{a}{b} - \frac{a'}{b'}| = \frac{|ab' - a'b|}{bb'} \geqslant \frac{1}{bb'}.$$

Now down to business. We first prove that convergents to $\alpha$ are record approximants in a certain sense.

**Theorem 1.** *Let $\alpha$ be an irrational real number and let $\frac{p_i}{q_i}$ be the convergents to $\alpha$. Suppose that $k$ is odd. Then any rational in the interval $(\frac{p_{k-1}}{q_{k-1}}, \frac{p_k}{q_k})$ has denominator greater than $q_k$. A similar statement holds when $k$ is even.*

*Proof.* Suppose that some rational $\frac{p}{q}$ in lowest terms lies in the interval stated and that $q \leqslant q_k$. We split into two cases.

*Case 1.* $\alpha < \frac{p}{q} < \frac{p_k}{q_k}$. Then

$$|\frac{p_k}{q_k} - \frac{p}{q}| \geqslant \frac{1}{qq_k} > \frac{1}{q_k q_{k+1}} > |\alpha - \frac{p_k}{q_k}|,$$

contradiction.

*Case 2.* $\frac{p_{k-1}}{q_{k-1}} < \frac{p}{q} < \alpha$. Then

$$|\frac{p_{k-1}}{q_{k-1}} - \frac{p}{q}| \geqslant \frac{1}{qq_{k-1}} \geqslant \frac{1}{q_k q_{k-1}} > |\alpha - \frac{p_{k-1}}{q_{k-1}}|,$$

contradiction. $\square$

**Lemma 1.** *For all $k$ we have $|\alpha - \frac{p_{k-1}}{q_{k-1}}| > |\alpha - \frac{p_k}{q_k}|$.*

*Proof.* Since $\frac{p_{k-1}}{q_{k-1}}$ and $\frac{p_k}{q_k}$ lie on opposite sides of $\alpha$, it is enough to show that $|\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k}| > 2|\alpha - \frac{p_k}{q_k}|$. However we have $|\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k}| \geqslant \frac{1}{q_{k-1}q_k}$ and $|\alpha - \frac{p_k}{q_k}| < \frac{1}{q_k q_{k+1}}$, and so it suffices to show that $q_{k+1} \geqslant 2q_{k-1}$. This follows from the fact that $q_{k+1} = a_{k+1}q_k + q_{k-1}$. $\square$

Putting these facts together, we immediately obtain the following result.

**Theorem 2.** *Suppose that $\alpha$ is an irrational number and that $\frac{p_k}{q_k}$ is a convergent to $\alpha$. Then the only rational $\frac{p}{q}$ with $q \leqslant q_k$ for which $|\alpha - \frac{p}{q}| \leqslant |\alpha - \frac{p_k}{q_k}|$ is $\frac{p_k}{q_k}$ itself.*

Secondly, we prove that all good approximants to $\alpha$ are convergents.

**Theorem 3.** *Suppose that $\frac{p}{q}$ is a fraction in lowest terms and that $|\alpha - \frac{p}{q}| \leqslant \frac{1}{2q^2}$. Then $\frac{p}{q} = \frac{p_k}{q_k}$ for some $k$.*

*Proof.* Since the denominators $q_k$ are increasing, we may select a unique $k$ such that $q_k \leqslant q < q_{k+1}$. Suppose that $\frac{p_k}{q_k} < \alpha < \frac{p_{k+1}}{q_{k+1}}$ (the other case is very similar).

First note that if $\frac{p}{q} < \frac{p_k}{q_k}$ then

$$|\alpha - \frac{p}{q}| > |\frac{p}{q} - \frac{p_k}{q_k}| \geqslant \frac{1}{qq_k} \geqslant \frac{1}{q^2},$$

contrary to assumption.

If $\frac{p}{q} = \frac{p_k}{q_k}$ then we are done. Suppose, then, that $\frac{p_k}{q_k} < \frac{p}{q}$. We have

$$|\frac{p}{q} - \frac{p_k}{q_k}| \geqslant \frac{1}{qq_k} > \frac{1}{q_k q_{k+1}} > |\alpha - \frac{p_k}{q_k}|,$$

so in fact $\frac{p_k}{q_k} < \alpha < \frac{p}{q}$. By the previous theorem, we cannot have $\frac{p}{q} \leqslant \frac{p_{k+1}}{q_{k+1}}$, and hence $\frac{p_k}{q_k} < \alpha < \frac{p_{k+1}}{q_{k+1}} < \frac{p}{q}$. We now divide into two cases.

*Case 1.* ($q$ large). Suppose that $q \geqslant \frac{1}{2}q_{k+1}$. Then

$$|\alpha - \frac{p}{q}| > |\frac{p}{q} - \frac{p_{k+1}}{q_{k+1}}| \geqslant \frac{1}{qq_{k+1}} \geqslant \frac{1}{2q^2},$$

contrary to assumption.

*Case 2.* ($q$ small). Suppose that $q < \frac{1}{2}q_{k+1}$. Then

$$|\alpha - \frac{p}{q}| = |\frac{p}{q} - \frac{p_k}{q_k}| - |\alpha - \frac{p_k}{q_k}| \geqslant \frac{1}{qq_k} - \frac{1}{q_k q_{k+1}} = \frac{1}{q_k}(\frac{1}{q} - \frac{1}{q_{k+1}}) > \frac{1}{2q^2},$$

also a contradiction.

$\square$

Continued Fractions

* Periodic Continued Fractions and Quadratic Irrationals

We saw that $\frac{1+\sqrt{5}}{2} = [1, 1, \dots]$ is periodic. Define a quadratic irrational to be an $\alpha$ which is not rational, but satisfies some quadratic equation $a\alpha^2 + b\alpha + c = 0$, $a, b, c \in \mathbb{Z}$. These are numbers of the form $\frac{A+B\sqrt{D}}{C}$, $A, B, C, D \in$ [illegible] $D > 0$.

Theorem

A number $\alpha$ has periodic continued fraction expansion (that is, the partial quotients $a_i$ repeat from some point on) $\Leftrightarrow$ $\alpha$ is a quadratic irrational.

Proof

We will first show the easier direction, namely that periodic continued fractions represent quadratic irrationals. We begin with the purely periodic case $\alpha = [a_0, a_1, \dots, a_k, a_0, a_1, \dots]$

Then $\alpha$ solves the equation

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{\alpha}}}$$

Rearranging, (subtract $a_0$, reciprocate, subtract $a_1$, etc) gives an equation of the form $\frac{c_1\alpha + c_2}{c_3\alpha + c_4} = \alpha$, $c_i \in \mathbb{Z}$

This is a quadratic equation for $\alpha$. Since $\alpha$ is uniquely determined by $a_0, a_1, \dots$ this equation cannot be trivial, and so $\alpha$ is a quadratic irrational.

This can be used in practice to evaluate any repeated continued fraction.

For the general case, suppose that $\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{\beta}}}$ with $\beta$ purely periodic. Now observe that $\beta$ is a quadratic irrational, and the set of quadratic irrationals is closed under reciprocating and addition of integers.

e.g. $\frac{C}{A+B\sqrt{D}} = \frac{C(A-B\sqrt{D})}{A^2-DB^2}$

The other direction is much trickier.

Suppose that $\alpha$ satisfies $a\alpha^2 + b\alpha + c = 0$, which we will write as $(\alpha \;\; 1)\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}\begin{pmatrix} \alpha \\ 1 \end{pmatrix} = 0$. Now we define $r_n$ by $\alpha = [a_0, a_1, a_2, \dots, a_{n-1}, r_n]$

Then, expanding out $[a_0, a_1, a_2, \dots, a_{n-1}, r_n]$, we get $\alpha = \frac{\tilde{p}_n}{\tilde{q}_n} = \frac{r_n p_{n-1} + p_{n-2}}{r_n q_{n-1} + q_{n-2}}$ from the recurrence relations for convergents.

Hence $\begin{pmatrix} \alpha \\ 1 \end{pmatrix} \propto \begin{pmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{pmatrix}\begin{pmatrix} r_n \\ 1 \end{pmatrix}$ (we write $\begin{pmatrix} x \\ y \end{pmatrix} \propto \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$)

Hence $(r_n \;\; 1)\begin{pmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{pmatrix}^T \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}\begin{pmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{pmatrix} = 0$

Hence $r_n$ satisfies the quadratic equation

$A_n r_n^2 + B_n r_n + C_n = 0$

where $A_n = a p_{n-1}^2 + b p_{n-1} q_{n-1} + c q_{n-1}$

$C_n = a p_{n-2}^2 + b p_{n-2} q_{n-2} + c q_{n-2}^2$

Taking determinants, $\begin{vmatrix} A_n & B_n/2 \\ B_n/2 & C_n \end{vmatrix} = \begin{vmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{vmatrix}^2 \begin{vmatrix} a & b/2 \\ b/2 & c \end{vmatrix}$

$= \begin{vmatrix} a & b/2 \\ b/2 & c \end{vmatrix}$, since $p_{n-1} q_{n-2} - p_{n-2} q_{n-1} = (-1)^n$

Clearly $A_n, B_n, C_n \in \mathbb{Z}$. We claim that $|A_n|, |C_n|$ are bounded independently of $n$. It then follows that $B_n$ is too (since $B_n^2 - 4A_nC_n = b^2 - 4ac$). Hence $r_n$ satisfies one of boundedly many equations. Hence $r_n = r_{n'}$ for some $n \neq n'$, at which point we are done.

We will show that $A_n$ is bounded ($C_n$ very similar).

$A_n = q_{n-1}^2\left(a\left(\frac{p_{n-1}}{q_{n-1}}\right)^2 + b\left(\frac{p_{n-1}}{q_{n-1}}\right) + c\right) = a\, q_{n-1}^2\left(\frac{p_{n-1}}{q_{n-1}} - \alpha\right)\left(\frac{p_{n-1}}{q_{n-1}} - \bar{\alpha}\right)$

where $\alpha, \bar{\alpha}$ are the roots of $ax^2 + bx + c = 0$.

But $\left|\frac{p_{n-1}}{q_{n-1}} - \alpha\right| \leq \frac{1}{q_{n-1}^2}$ since $\frac{p_{n-1}}{q_{n-1}}$ is a convergent.

Also, $\left|\frac{p_{n-1}}{q_{n-1}} - \bar{\alpha}\right| \leq |\alpha| + 1 + |\bar{\alpha}|$ if $n$ is large enough that $\left|\frac{p_{n-1}}{q_{n-1}} - \alpha\right| \leq 1$.

Putting this together gives $|A_n| \leq \alpha\left(|\alpha| + 1 + |\bar{\alpha}|\right)$

* The Gauss Map and Continued Fraction Expansions of Typical $\alpha$

Remember that $\alpha = [a_0, a_1, \dots, a_{k-1}, a_k + \varepsilon_k]$ and we have $\varepsilon_k = \frac{1}{a_{k+1} + \varepsilon_{k+1}}$, with $\varepsilon_{k+1} = T(\varepsilon_k)$, where $T(x) = \left\{\frac{1}{x}\right\}$ is the Gauss Map.

Question

What is the 'probability' that $\varepsilon_{k+1} \in [0, c]$ for some $c$?

One might expect

i) The probability that $\varepsilon_k \in [0, c]$ is the same as for $\varepsilon_{k+1}$

ii) The probability is given by a formula

$$\mathbb{P}\left(\varepsilon_k \in [0, c]\right) = \int_0^c f(x)\,dx$$

3

## Continued Fractions and the Gauss Map

Recall $\alpha = [a_0, a_1, \dots, a_{k-1}, a_k, \varepsilon_k]$, $T(\varepsilon_k) = \varepsilon_{k+1}$

$T(x) = \{\frac{1}{x}\}$. For random $\alpha \in [0,1]$, one imagines that $\mathbb{P}(\varepsilon_k \in [0,c])$ should be independent of $k$ and should be "nice", $\mathbb{P}(\varepsilon_k \in [0,c]) = \int_0^c f(x)\,dx$. This is a reasonable thing to imagine, akin to decimal expansion, where $\mathbb{P}(k^{th} \text{ digit} = m) = \frac{1}{10}$. The analog of $\varepsilon_k$ is the base 10 expansion after the $k^{th}$ digit, $0\cdot 00\ldots 0\, d_k d_{k+1}\ldots$. The remainder $0\cdot d_k d_{k+1}\ldots$ is uniformly distributed in $[0,1]$.

<u>What is f</u>?

We should have $\mathbb{P}(\varepsilon_{k+1} \in [0,c]) = \int_0^c f(x)\,dx$

but also $= \mathbb{P}\left(\varepsilon_k \in \bigcup_{j\geq 1}\left[\frac{1}{j+c}, \frac{1}{j}\right]\right) = \sum_{j\geq 1} \int_{\frac{1}{j+c}}^{\frac{1}{j}} f(x)\,dx.$

since $T^{-1}([0,c]) = \bigcup_{j\geq 1}\left[\frac{1}{j+c}, \frac{1}{j}\right]$

It has been observed that $f(x) = \frac{1}{1+x}$ satisfies this equation. $\int_0^c f(x)\,dx = \log(1+c)$

$$\sum_{j\geq 1} \int_{\frac{1}{j+c}}^{\frac{1}{j}} f(x)\,dx = \sum_{j\geq 1} \log\left(1+\tfrac{1}{j}\right) - \log\left(1+\tfrac{1}{j+c}\right)$$

$$= \lim_{m\to\infty}\left(\sum_{j=1}^{m}(\log(j+1) - \log j) - \sum_{j=1}^{m}(\log(1+j+c) - \log(j+c))\right)$$

$$= \lim_{m\to\infty}\left(\log(m+1) - \log(1+m+c)\right) + \log(1+c)$$

But $\lim_{m\to\infty}\left(\log(m+1) - \log(1+m+c)\right) = 0$

To make sure that $\int_0^1 f(x)\,dx = 1$, we take $f(x) = \frac{1}{\log 2}\frac{1}{1+x}$. The measure defined by $\mu(A) = \int_A f(x)\,dx$ is called the <u>Gauss Measure</u>.

This "should be" the probability that $\varepsilon_k$ lies in $A$.

** To put this on a rigorous footing, proceed as follows. The Gauss measure is <u>invariant</u> for the Gauss Map, $\mu(T^{-1}(A)) = \mu(A)$. We proved this when $A = [0, c]$ and the general case follows by taking limits (ie. when $A$ is Lebesgue measurable). Key fact: $T$ is ergodic for this measure.

<u>Ergodic</u> means that $T$ has no non-trivial invariant sets i.e. $T^{-1}(A) = A \Rightarrow \mu(A) = 0$ or $1$.

<u>Ergodic Theorem</u>

"Time Averages = Space Averages"

$$\lim_{N \to \infty} \frac{1}{N}\left(\Psi(x) + \Psi(T(x)) + \dots + \Psi(T^{N-1}(x))\right) = \frac{1}{\log 2}\int_0^1 \frac{\Psi(t)}{1+t}\,dt$$

for almost all $x$ and "nice" $\Psi$.

In other words, for almost all $x$, $x, T(x), T^2(x), \dots$ are distributed according to the Gauss measure. **

What is the probability that the $k^{th}$ partial quotient of a random $\alpha$ is equal to $m$?

This is
$$\mathbb{P}\left(\varepsilon_k \in \left(\tfrac{1}{m+1}, \tfrac{1}{m}\right]\right) = \frac{1}{\log 2}\int_{\frac{1}{m+1}}^{\frac{1}{m}} \frac{dx}{1+x}$$
$$= \frac{1}{\log 2}\left(\log\left(1+\tfrac{1}{m}\right) - \log\left(1+\tfrac{1}{m+1}\right)\right) = \frac{1}{\log 2}\log\left(1+\tfrac{1}{m(m+2)}\right) \asymp \frac{c}{m}$$

(Hence the fact that 292 is a partial quotient of $\pi$, meaning that $\frac{355}{113}$ is such a good approximation to $\pi$, has "probability about 1%")

* <u>Continued Fraction of $e$</u>

$$e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \dots]$$
$$= [1, 0, 1, 1, 2, 1, 1, 4, \dots]$$

Here is a sketch proof:

If $p_n/q_n$ are the ~~partial~~ convergents, since $a_{3n}=1$, $a_{3n+1}=2n$, $a_{3n+2}=1$, we should have

$$p_{3n}=p_{3n-1}+p_{3n-2},\quad p_{3n+1}=2np_{3n}+p_{3n-1},\quad p_{3n+2}=p_{3n+1}+p_{3n}$$

and similar for $q_n$.

Now consider $A_n=-\int_0^1 \frac{t^n(1-t)^n}{n!}e^t dt$

$$B_n=\int_0^1 \frac{t^{n+1}(1-t)^n}{n!}e^t dt,\quad C_n=\int_0^1 \frac{t^n(1-t)^{n+1}}{n!}e^t dt$$

Claim

$$p_{3n}-eq_{3n}=A_n,\quad p_{3n+1}-eq_{3n+1}=B_n,\quad p_{3n+2}-eq_{3n+2}=C_n$$

Proof by Induction

Given the recurrence relations, it is enough to check $n=0$ (easy), and (1) $A_n=C_{n-1}+B_{n-1}$, (2) $B_n=2nA_n+C_{n-1}$

$$C_n=A_n+B_n$$

To prove (1) look at $\int_0^1 \frac{d}{dt}\left(t^n(1-t)^n e^t\right) dt = 0$

To prove (2) look at $\int_0^1 \frac{d}{dt}\left(t^n(1-t)^{n+1} e^t\right) dt = 0$

But $|A_n|, |B_n|, |C_n| \le e/n!$, so it follows that

$$\frac{p_{3n}}{q_{3n}},\ \frac{p_{3n+1}}{q_{3n+1}},\ \frac{p_{3n+2}}{q_{3n+2}} \to e.$$

Corollary

$e^2$ is irrational. Indeed, if $e^2$ were rational, the continued fraction expansion of $e$ would be periodic.

Open Problems

1. Does $\sqrt[3]{2}$ have bounded partial quotients?
2. (Littlewood Conjecture) Suppose that $\alpha, \beta$ are independent

Are there arbitrarily large $q_n$ such that $q_n \{\alpha\}\{\beta\} \to 0$

Note that if the $q_n$ are the denominators of convergents to $\alpha$, then $|\alpha - p/q| \leq \frac{1}{q^2} \Rightarrow q\{\alpha\} \leq 1$, and similarly for $\beta$. But how to do this simultaneously?

Transcendental Numbers

Definition

$\alpha$ is algebraic if it satisfies a non-trivial polynomial equation

$a_n \alpha^n + a_{n-1} \alpha^{n-1} + \dots + a_1 \alpha + a_0 = 0$, $a_i \in \mathbb{Z}\ \forall i$

e.g. $\alpha = \sqrt{2}, \sqrt{3}, \sqrt{5} + \sqrt{7}, 10^{1/7}$ etc

If $\alpha$ is not algebraic, it is transcendental. Cantor:

There exist transcendental numbers (Numbers and Sets).

# Transcendental Numbers

The key idea is that if $\alpha$ is algebraic then $\alpha$ is either rational or is not "too well approximated" by rationals.

**Theorem** (Liouville)

Suppose that $\alpha$ satisfies a polynomial $a_n\alpha^n + \ldots + a_0 = 0$ with $a_i \in \mathbb{Z}$. Suppose that $\alpha$ is not rational. Then there is a constant $c(\alpha)$ such that $|\alpha - p/q| \geq \frac{c(\alpha)}{q^n}$ for all integers $p, q$, $p, q \neq 0$.

**Proof**

WLOG $p/q$ is not a root of the equation satisfied by $\alpha$ (if it is, divide by $X - p/q$ and clear denominators).

Since an integer that is not zero must have magnitude at least 1, $|a_n(\frac{p}{q})^n + \ldots + a_0| \geq \frac{1}{q^n}$ since the left hand side is a non-zero integer over $q^n$. Comparing with $a_n\alpha^n + \ldots + a_0 = 0$, we get $|a_n(\alpha^n - (\frac{p}{q})^n) + \ldots + a_0| \geq \frac{1}{q^n}$

$$\Rightarrow |a_n||\alpha^n - (\tfrac{p}{q})^n| + \ldots + |a_0| \geq \frac{1}{q^n} \qquad (*)$$

But $|\alpha^m - (\frac{p}{q})^m| = |\alpha - p/q||\alpha^{m-1} + \alpha^{m-2}(\frac{p}{q}) + \ldots + (\frac{p}{q})^{m-1}|$

If $|\alpha - p/q| \geq 1$ then the theorem is immediate. Otherwise, the preceding is bounded by $|\alpha - p/q|\, c(m, \alpha)$, where $c(m, \alpha) = |\alpha|^{m-1} + |\alpha|^{m-2}(|\alpha|+1) + \ldots + (|\alpha|+1)^{m-1}$

Substituting back into $(*)$, we get

$$|\alpha - p/q|\, \hat{c}(\alpha) \geq 1/q^n$$

where $\hat{c}(\alpha) = |a_n|\, c(n, \alpha) + |a_{n-1}|\, c(n-1, \alpha) + \ldots + |a_0|$ $\square$

To make a transcendental number, one need only write down a non-rational that is very well approximated by rationals, such as $\alpha = \sum_j 10^{-j!}$

Theorem

$\alpha = \sum_j 10^{-j!}$ is transcendental.

Proof

Let $q = 10^{N!}$. Then $\alpha$ is equal to some rational $p/q$ plus some error $\sum_{j \geq N+1} 10^{-j!}$ which is at most $2 \cdot 10^{-(N+1)!}$, which is $2q^{-(N+1)}$. For any fixed $n$, $C$, this is smaller than $C/q^n$ if $N$, and hence $q$, are sufficiently large. Hence $\alpha$ is not the root of a polynomial of degree $n$, for any $n$ ↯

Remarks

1. We found a sufficient condition (being exceptionally well approximated by rationals) for being transcendental. It was not necessary. We will see in a moment that $e$ is transcendental. However, $e$ is not incredibly well approximated by rationals. In fact, $|e - p/q| \geq \frac{C_\varepsilon}{q^{2+\varepsilon}}$ for every $\varepsilon > 0$, which follows from the continued fraction expansion of $e$.
2. Roth famously proved the following improvement of Liouville's result: If $\alpha$ is algebraic and not rational, then for every $\varepsilon > 0$, there is a constant $C = C(\alpha, \varepsilon)$ such that $|\alpha - p/q| \geq \frac{C(\alpha,\varepsilon)}{q^{2+\varepsilon}}$. But unfortunately, $C(\alpha, \varepsilon)$ is not effectively computable.

Theorem

$e$ is transcendental. We will consider integrals of the form $I(t) = \int_0^t e^{t-u} f(u)\,du$ where $f$ is a polynomial.

Proof

Integrating by parts gives $I(t) = e^t f(0) - f(t) + \int_0^t e^{t-u} f'(u)\,d$

Doing this $D+1$ times, where $D = \deg(f)$, gives

$I(t) = e^t \sum_{j=0}^{D} f^{(j)}(0) - \sum_{j=0}^{D} f^{(j)}(t)$ (*)

Suppose for contradiction that $b_0 + b_1 e + \dots + b_r e^r = 0$ with $b_0, b_1, \dots, b_r \in \mathbb{Z}$, $b_r \neq 0$. Let $p$ be a large prime and consider $f(x) = x^{p-1}(x-1)^p \dots (x-r)^p$ (the clever bit)

Applying (*) with this $f$, we get

$\sum_{i=0}^{r} b_i I(i) = -\sum_{i=0}^{r} b_i \sum_{j=0}^{D} f^{(j)}(i)$ $(D = \deg(f) = rp + p - 1,$

Key observations

1. The left hand side behaves like $C^p$ as $p \to \infty$
2. Everything on the right hand side is divisible by $p!$ except for one term only divisible by $(p-1)!$ $\Rightarrow$ Right Hand Side $\gtrsim (p-1)!$

Details

1. We have $|f(x)| \leq (r^{r+1})^p$ for $0 \leq x \leq r$ (very crude). Hence $|I(t)| \leq r(e^r r^{r+1})^p$ for $0 \leq t \leq r$.

   Thus, $\sum_{i=0}^{r} b_i I(i) \leq r(r+1) \max_i |b_i| (e^r r^{r+1})^p \leq (2e^r r^{r+1})^p$ if $p$ is large enough.
2. If we differentiate $f(x) = x^{p-1}(x-1)^p \dots (x-r)^p$ less than $p$ times, then set $x = 0, 1, \dots, r$, we will always

get $0$ except that $f^{(p-1)}(0) = (p-1)!\,((-1)^r r!)^p$ (product rule

On the other hand, if we differentiate $p$ or more times, $f^{(j)}(i)$ is always divisible by $p!$. Hence

$$RHS = -b_0 (p-1)!\,((-1)^r r!)^p + \text{multiple of } p!.$$

Choose $p$ so large that $p \nmid b_0$ and $p \nmid r!$. Then $RHS \geq (p-1)!$. But for $p$ large enough, $(p-1)! > C^p$, for any $p$. $\square$

The Baire Category Theorem

Let $X$ be a metric space. We will be talking about complete metric spaces - every Cauchy sequence has a limit.

If $A \subseteq X$, then we say that $A$ is dense if $A$ intersects every open ball $B_\varepsilon(x)$ in $X$.

e.g. the rationals, $\mathbb{Q}$, and irrationals, $\mathbb{R} \setminus \mathbb{Q}$, are dense in $\mathbb{R}$

Theorem (Baire-Category Theorem)

Let $X$ be a non-empty complete metric space. Let $(A_n)_{n=1}^{\infty}$ be a sequence of dense open sets. Then $\bigcap_{n=1}^{\infty} A_n$ is non-empty (and in fact dense).

Equivalent Formulation

Let $X$ be a non-empty complete metric space. Suppose that $X = \bigcup_{n=1}^{\infty} F_n$ with each $F_n$ closed. Then one of these sets $F_n$ has non-empty interior; that is to say that it contains some $B_\varepsilon(x)$, $\varepsilon > 0$.

"A complete metric space is not a countable union of $\overbrace{\text{small closed sets}}^{\text{meagre}}$."

Proof

We will repeatedly use the fact that $B_\varepsilon(x) \supseteq \overline{B_{\varepsilon/2}(x)}$

Pick a ball $B_{\varepsilon_0}(x_0)$ in $X$. This intersects $A_1$ (since $A_1$ is dense). Since $A_1$ is open, $A_1 \cap B_{\varepsilon_0}(x_0)$ contains an open ball, and hence, contains $\overline{B_{\varepsilon_1}(x_1)}$ of some open ball. Since $A_2$ is dense, it intersects $B_{\varepsilon_1}(x_1)$, and the intersection contains an open ball, and hence the closure $\overline{B_{\varepsilon_2}(x_2)}$ of some open ball.

We continue in this way to produce a nested sequence $B_{\varepsilon_i}(x_i)$ such that $\overline{B_{\varepsilon_i}(x_i)} \subseteq A_1, \ldots, A_i$, $B_{\varepsilon_{i+1}}(x_{i+1}) \subseteq B_{\varepsilon_i}(x_i) \cap A_{i+1}$

Make sure that the $\varepsilon_i$ are chosen so that $\varepsilon_i \to 0$. Then the centre of these balls, $x_i$, form a Cauchy sequence. Thus, there is some $x$ such that $x_i \to x$. We must have $x \in \overline{B_{\varepsilon_i}(x_i)}$ and hence $x \in \bigcap_{n=1}^{\infty} A_n$. To see that $\bigcap_{n=1}^{\infty} A_n$ is in fact dense, replace $X$ by $X' = \overline{B_\delta(x)}$ for any $x \in X$, $\delta > 0$. $X'$ is a complete metric space, and the sets $A_n' = A_n \cap X'$ are open in the subspace topology on $X'$ and are dense. Thus $\bigcap A_n'$ is non-empty and lies in $\overline{B_\delta(x)}$ □

Proposition

There exists a continuous, nowhere differentiable function $f : [0,1] \to \mathbb{R}$

(Such a function can be constructed explicitly)

Proof

Let $X = C[0,1]$ be the metric space of continuous functions on $[0,1]$ together with the supremum norm.

This is complete (uniform limits of continuous functions are continuous).

We will define a sequence $(A_n)_{n=1}^{\infty}$ of closed $A_n \subseteq X$ such that

i) If $f$ is differentiable at some point of $(0,1)$, then $f$ lies in some $A_n$

ii) Each $A_n$ has empty interior.

The Baire Category Theorem then implies that $X \neq \bigcup_{n=1}^{\infty} A_n$. Then any $f \in X \setminus \bigcup_{n=1}^{\infty} A_n$ is continuous but nowhere differentiable.

Define $A_n$ to be the set of $f \in X$ such that there exists $x \in [0,1]$ with the property that $\left|\frac{f(x)-f(y)}{x-y}\right| \leq n$ whenever $0 < |x-y| \leq \frac{1}{n}$. Note that if $f$ is differentiable at $x$ then $\lim_{y \to x} \frac{f(y)-f(x)}{y-x} = f'(x)$ exists. Thus if $n$ is sufficiently large, $f \in A_n$. This is i).

Why is $A_n$ closed?

Suppose that $f_i$ is a sequence of functions in $A_n$ and that $f_i \to f$. We need to show that $f \in A_n$. We know that for each $i$, there is an $x_i$ such that $\left|\frac{f_i(y)-f_i(x_i)}{y-x_i}\right| \leq n$ whenever $0 < |y-x_i| \leq \frac{1}{n}$. By Bolzano-Weierstrass, there is a subsequence of the $x_i$ converging to some point $x$. Relabelling, we assume that in fact, $x_i \to x$. Suppose that $0 < |y-x| < \frac{1}{n}$. Then if $i$ is large enough, we have $0 < |y-x_i| < \frac{1}{n}$ and so $\left|\frac{f_i(y)-f_i(x_i)}{y-x_i}\right| \leq n$. Now let $i \to \infty$. Since $f_i \to f$ uniformly and each $f_i$ is continuous, we get

$\lim_{i\to\infty} f_i(y) = f_i(y)$, $\lim_{i\to\infty} f_i(x_i) = f(x)$, and hence

$$\left|\frac{f(y)-f(x)}{y-x}\right| \leq n$$

Since $f$ is continuous, the same bound holds when $|y-x| = \frac{1}{n}$ as well. Hence $A_n$ is closed.

To complete the proof, we need to show that $A_n$ has empty interior

"Close to $f$ is any function with horribly unbounded slope."

We do this in two steps. Let $\varepsilon > 0$. Then

i) $f$ is within $\frac{\varepsilon}{2}$ (in the uniform norm) of a piecewise linear function $g$ (we can subdivide $[0, 1)$ into finitely many segments $[x_i, x_{i+1})$ on which $g$ is linear).

ii) Within $\frac{\varepsilon}{2}$ of any piecewise linear function is a function not in $A_n$.

i), ii) combine to show that within $\varepsilon$ of $f$, i.e. in $B_\varepsilon(f)$, there is a continuous function not in $A_n$. Thus $A_n$ has empty interior.

<u>Proof of i)</u>

$f$ is uniformly continuous, hence if $\delta$ is small enough, then $|f(x) - f(y)| < \varepsilon/4$ whenever $|x-y| \leq \delta$. Partition $[0, 1]$ into finitely many intervals $[x_i, x_{i+1})$ of width $\leq \delta$. Define $g : [0,1] \to \mathbb{R}$ by setting $g(x_i) = f(x_i)$ for all $i$ and taking $g$ linear on $[x_i, x_{i+1}$

It follows that $\|f - g\| \leq \varepsilon/2$.

## Baire Category Theorem and Pathological Functions

If every continuous function was differentiable at some point, then we would have $X = \bigcup_{n=1}^{\infty} A_n$. Baire $\Rightarrow$ some $A_n$ has non-empty interior. Suppose that $f \in X$ and $\varepsilon > 0$. We found a piecewise linear function $g$ with $\|f - g\| \leq \varepsilon/2$.


$\Psi$.

Now we find $h$ with $\|g - h\| \leq \varepsilon/2$ and with $h$ having slope $> n$ everywhere. Consider $\Psi$ as drawn, where $M$ is a large +ve integer and $\Psi(j/m) = (-1)^{j+1}$, $\Psi$ linear on each interval $[j/m, \frac{j+1}{m}]$

$\Psi$ is continuous, and it has slope $\geq 2m$ at every point in $[0,1]$. Define $h = g + \frac{\varepsilon}{2}\Psi$. Clearly $\|g - h\| \leq \varepsilon/2$. ~~is bounded~~ The slope of $g$ is bounded by some $C$ since $g$ is piecewise linear. The slope of $h$ is always at least $\frac{\varepsilon}{2} \cdot 2m - C$. Choosing $n$ large enough, this can be made $> n$. Thus, we have found $h \notin A_n$, $\|f - h\| \leq \varepsilon$. Since $\varepsilon$ was arbitrary, $A_n$ has empty interior. ⋇

## Uniform Boundedness Principle

### Theorem

Suppose that $X$ is a complete metric space. Suppose that $\mathcal{F}$ is a collection of continuous functions on $X$. Suppose that

$\forall x \in X$, $\sup_{f \in \mathcal{F}} |f(x)| < \infty$.

Then there is some ball $B_\delta(x_0)$, $\delta > 0$, such that

$\sup_{f \in \mathcal{F}} \sup_{x \in B_{x_0}(\delta)} |f(x)| < \infty$. (When $X = [0,1]$, this states that if $|f(x)| \leq C_x$, for all $f \in \mathcal{F}$, for some $C_x$, then the functions $f$ are in fact uniformly bounded on some interval $(a,b)$.

Proof

For every $n$, the set $\overbrace{\bigcap_{f\in\mathcal{F}} \{x\in X : |f(x)|\le n\}}^{=A_n}$ is closed.

Indeed, since $f$ is continuous, $f^{-1}([-n,n]) = \{x\in X : |f(x)|\le n\}$ is closed, and arbitrary intersections of closed sets are closed.

By assumption, $X = \bigcup_n A_n$. Thus by Baire, one of the $A_n$ has non-empty interior. $\square$

Points of Continuity

Theorem

or any countable, dense set

Suppose $f:[0,1]\to\mathbb{R}$ is continuous at every rational. Then $f$ is in fact continuous at uncountably many points of $[0,1]$.

Proof

The key claim is that the set of points where a function is continuous is a $G_\delta$-set, by which we mean a countable intersection of open sets.

variation of $f$ over $[x-\delta, x+\delta]$

To prove this, consider $\omega_f(x) = \lim_{\delta\to 0} \sup_{\substack{|y-x|\le\delta \\ |y'-x|\le\delta}} |f(y)-f(y')|$

The limit exists since the 'variation' quantity is non-increasing in $\delta$. We claim that $\omega_f(x) = 0 \iff f$ is continuous at $x$, an easy exercise. Furthermore, $\{x : \omega_f(x) < \varepsilon\}$ is open for all $\varepsilon > 0$. Indeed, suppose that $\omega_f(x) < \varepsilon$.

If $\delta$ is small enough, $\sup_{\substack{|y-x|\le\delta \\ |y'-x|\le\delta}} |f(y)-f(y')| \le \varepsilon' < \varepsilon$.

If $|x'-x| \le \delta/2$ then

$$\sup_{\substack{|y-x'|\le\delta/2 \\ |y'-x'|\le\delta/2}} |f(y)-f(y')| \le \sup_{\substack{|y-x|\le\delta \\ |y'-x|\le\delta}} |f(y)-f(y')| \le \varepsilon' < \varepsilon.$$

Hence, the set of points of continuity of $f$ is $\bigcap_{n\geq 1}\{x : \omega_f(x) < \frac{1}{n}\}$ is a $G_\delta$ set.

Claim

$Q = \mathbb{Q} \cap [0,1]$, the rationals in $[0,1]$, is not a $G_\delta$-set.

Proof

Suppose $Q = \bigcap_{n=1}^{\infty} U_n$ $U_n \subseteq [0,1]$ open.

Then $[0,1] = \bigcup_{q\in Q}\{q\} \cup \bigcup_{n=1}^{\infty} U_n^c$. This is a countable union of closed sets. Thus, by Baire, one of these has non-empty interior. Clearly $\{q\}$ has empty interior. Alternatively, $U_n^c$ contains an open ball. But this cannot happen, since $\mathbb{Q}$ intersects every open ball. $\square$

Banach-Tarski "Paradox"

Theorem

Assume the Axiom of Choice. Then, we may divide the unit ball $X$ in $\mathbb{R}^3$ into 17 pieces $A_1, \dots, A_{17}$ which may be rearranged to give two copies of $X$. More precisely, there are isometries $g_1, \dots, g_{17} \in \mathrm{Isom}^+(\mathbb{R}^3)$ (rotations and translations, orientation preserving) such that $g_1A_1 \cup \dots \cup g_{17}A_{17}$ is two disjoint copies of $X$.

Remark

There is no such example for $\mathbb{R}$ or $\mathbb{R}^2$.

Why is this not really a paradox? There is no notion of "weight" or volume valid for all subsets $A \subseteq X$. More mathematically, there does not exist a finitely additive measure $\mu : \mathcal{P}X \to [0, \infty)$ such that

i) $\mu(gA) = \mu(A)$ for any isometry $g$.

ii) $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A \cap B = \emptyset$

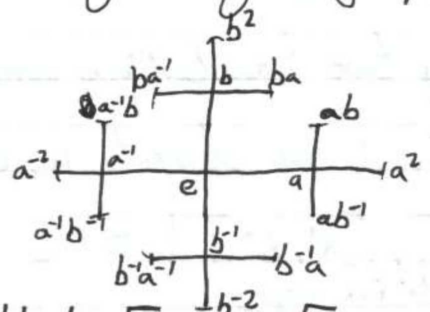
iii) $\mu(X) = \frac{4}{3}\pi$

Topics in Analysis

## The Free Group on Two Generators

This group consists of generators $a, b$, and all words in $a, b$, $a^{-1}, b^{-1}$ such that no word is equal to the identity unless it obviously is (e.g. $ab^{-1}ba^{-1}$ is obviously equal to the identity). More formally, no <u>reduced word</u> such as $a^4b^{-7}a^6$ is equal to the identity. For example, $a^3b^6b^{-2}a^5a^{-3}b$ is not reduced, but reduces to $a^3b^4a^2b$.

A picture of the free group, $F$, on two generators:




Write $F_a$ for the words beginning with $a$. Fairly obviously,

$$F = F_a \cup F_b \cup F_{a^{-1}} \cup F_{b^{-1}} \cup \{e\}$$

Note that $F_a \cup aF_{a^{-1}}$ is the whole of $F$, as is $F_b \cup bF_{b^{-1}}$

Thus, the free group $F$ can be decomposed into 4 pieces (plus $e$) which can be rearranged to give two copies of $F$.

Key idea in Banach-Tarski: One can decompose the ball $B^3 \subseteq \mathbb{R}^3$ into tree-like structures. To do this, we will show that $\mathrm{Isom}^+(\mathbb{R}^3)$ (and in fact the group $SO(3)$ of rotations in $\mathbb{R}^3$) contains a free subgroup.

Why is this not the case in $\mathbb{R}$ or $\mathbb{R}^2$?

$\mathrm{Isom}^+(\mathbb{R}) \cong \mathbb{R}$ (translations). This group is <u>abelian</u>, and hence doesn't contain the free group $F$, since $aba^{-1}b^{-1} = e$ always.

$\mathrm{Isom}^+(\mathbb{R}^2)$ is generated by the group $SO(2)$ of rotations

about $O$ together with translations. Every such isometry has the form $x \mapsto Ax + b$, $A \in SO(2)$, $b \in \mathbb{R}^2$.
This gives $Isom^+(\mathbb{R}^2) \cong SO(2) \rtimes \mathbb{R}^2$ with
$(A, b) * (A', b') = (A'A, A'b + b')$.
This group is not abelian. Indeed, if $g_1 = (A, b)$, and $g_2 = (A', b')$, then $[g_1, g_2] = g_1 g_2 g_1^{-1} g_2^{-1}$
$= (A'^{-1} A^{-1} A' A, *) = (I, *)$, a translation, since $SO(2)$ is abelian. Hence $[[g_1, g_2], [g_3, g_4]] = e$ for all $g_1, \dots, g_4$, since any two translations commute.
Hence if $Isom^+(\mathbb{R}^2)$ contained two elements $a, b$ generating a free group, we have $\underbrace{[[a,b],[a^2,b^2]]}_{\text{not the trivial word}} = e$
This is a contradiction ($Isom^+(\mathbb{R}^2)$ is a soluble group).

<u>Theorem</u>

$SO(3)$ contains two elements $a, b$ generating a free group.

<u>Proof</u>

$$a = \begin{pmatrix} \frac{1}{3} & -\frac{2\sqrt{2}}{3} & 0 \\ \frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2\sqrt{2}}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix}$$

rotations through $\arccos(\frac{1}{3})$ about the $z$ and $x$ axes respectively.
Suppose that there is a reduced word $w$ which is the identity.
By considering $a^{-m} w a^m$ for large $m$, we assume that the word $w$ ends in $a$.
Write $w = g_n \cdots g_1$, $g_i \in \{a, a^{-1}, b, b^{-1}\}$ and $g_1 = a$. We will show that $w\begin{pmatrix}1\\0\\0\end{pmatrix} \neq w\begin{pmatrix}1\\0\\0\end{pmatrix}$ and derive a contradiction.

Topics in Analysis

We have $g_n \dots g_1 \begin{pmatrix}1\\0\\0\end{pmatrix} = 3^{-n}\begin{pmatrix}x_n\\ y_n\sqrt{2}\\ z_n\end{pmatrix}$ where $x_n, y_n, z_n \in \mathbb{Z}$.

By induction:

$$(x_{n+1}, y_{n+1}, z_{n+1}) = \begin{cases} (x_n - 4y_n,\ 2x_n + y_n,\ 3z_n), & g_{n+1} = a \\ (x_n + 4y_n,\ -2x_n + y_n,\ 3z_n), & g_{n+1} = a^{-1} \\ (3x_n,\ y_n + 2z_n,\ -4y_n + z_n), & g_{n+1} = b \\ (3x_n,\ y_n - 2z_n,\ 4y_n + z_n), & g_{n+1} = b^{-1} \end{cases}$$

Write $(x_{n+1}, y_{n+1}, z_{n+1}) = \Phi_{g_{n+1}}(x_n, y_n, z_n)$

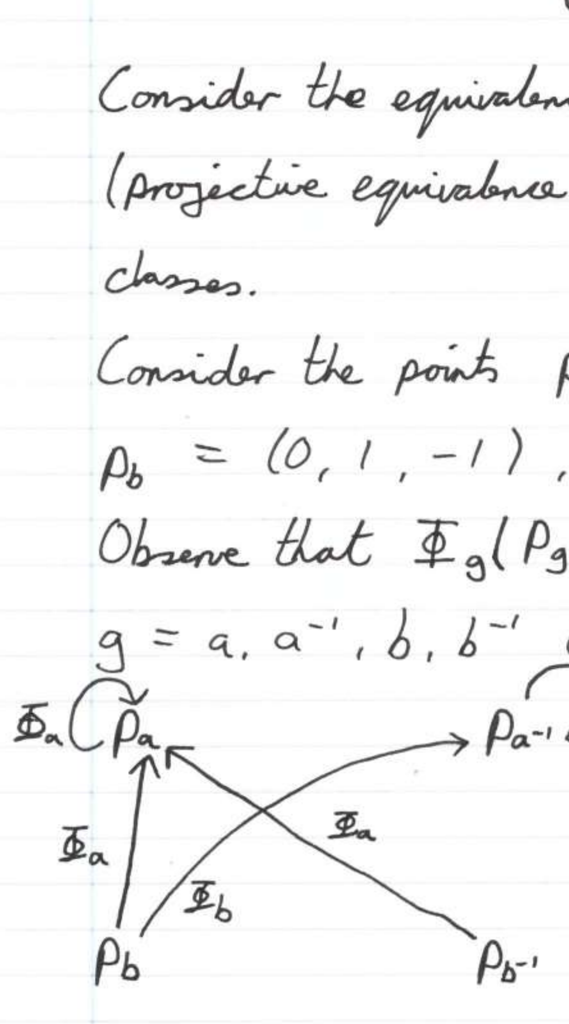
Now consider the vectors $(x, y, z) \pmod 3$. Let $\bar{\Phi}_g$ be the corresponding map.

$$\bar{\Phi}_g(x, y, z) = \begin{cases} (x - y,\ -x + y,\ 0), & g = a \\ (x + y,\ x + y,\ 0), & g = a^{-1} \\ (0,\ y - z,\ -y + z), & g = b \\ (0,\ y + z,\ y + z), & g = b^{-1} \end{cases}$$

Consider the equivalence relation $(x, y, z) \sim (-x, -y, -z)$ (projective equivalence). The maps $\bar{\Phi}_g$ act on these equivalence classes.

Consider the points $p_a = (-1, 1, 0)$, $p_{a^{-1}} = (1, 1, 0)$, $p_b = (0, 1, -1)$, $p_{b^{-1}} = (0, 1, 1)$ in this space.

Observe that $\bar{\Phi}_g(p_{g'}) = p_g$ unless $g' = g^{-1}$ for $g = a, a^{-1}, b, b^{-1}$ (proof by inspection).




"$p_g$ is an attracting fixed point for $\bar{\Phi}_g$".

Note that $\Phi_a \begin{pmatrix}1\\0\\0\end{pmatrix} = p_a$.

Hence if $g_n \cdots g_1$ is a reduced word ending in $a$, $\Phi_{g_n} \Phi_{g_{n-1}} \cdots \Phi_{g_1} \begin{pmatrix}1\\0\\0\end{pmatrix}$ bounces around the diagram, but never ends up back at $\begin{pmatrix}1\\0\\0\end{pmatrix}$

"Proof by ping-pong"

Theorem

There is a free subgroup on two (quite explicit) generators in $SO(3)$

Remarks

i) There is actually a copy of $F_n$, the free subgroup on $n$ generators in $SO(3)$. In fact (not quite trivial) $F_2$ contains a copy of $F_n$.

ii) A random pair of elements in $SO(3)$ (whatever this means) will generate a free group. "Proof": Consider a word $w(a,b)$ such as $a^3b^7a^{-11}$. The set $w(a,b) = id$ is a proper subvariety of $SO(3)$, given by a polynomial. This is proper since there is a choice $a = a_*$, $b = b_*$ for which $w(a_*, b_*) = id$ as we saw last time. There are countably many words, and for each one, $\mathbb{P}(w(a,b) = id) = 0$. Hence $\mathbb{P}(w(a,b) = id \text{ for some } w) = 0$.

Hausdorff Paradox ("Baby" Banach-Tarski)

Theorem

There is a countable set $D \subset S^2 \subset \mathbb{R}^3$ such that $X = S^2 \setminus D$ has the following property: $\exists$ disjoint $A_1, \dots, A_4 \subset X$, $g_1, \dots, g_4 \in SO(3)$ such that $g_1A_1 \cup g_2A_2 = g_3A_3 \cup g_4A_4 = X$

Proof

Let $F$ be a free group on the generators $a, b$ in $SO(3)$. $F$ is countable. Let $D$ consist of the two (antipodal) fixed points of all rotations in $F$. Let $X = S^2 \setminus D$. Then $F$ acts on $X$ by rotations. $X$ splits up into orbits $\{gx : g \in F\}$ under this action.

Let $Y \subset X$ be a set containing precisely one point from each orbit. (Serious use of AC). Then $X$ is the disjoint union $\bigcup_{y \in Y} \bigcup_{g \in F} g \cdot y$. Recall from last time that we may split $F = F_a \cup F_{a^{-1}} \cup F_b \cup F_{b^{-1}} \cup \{id\}$ such that $F = F_a \cup aF_{a^{-1}} = F_b \cup bF_{b^{-1}}$

Now define $A_1 = \bigcup_{y \in Y} \bigcup_{g \in F_a} g \cdot y \qquad A_3 = \bigcup_{y \in Y} \bigcup_{g \in F_b} g \cdot y$

$A_2 = \bigcup_{y \in Y} \bigcup_{g \in F_{a^{-1}}} g \cdot y \qquad A_4 = \bigcup_{y \in Y} \bigcup_{g \in F_{b^{-1}}} g \cdot y$

Take $g_1 = g_3 = id$, $g_2 = a$; $g_4 = b$ $\square$

<u>Proposition</u>

There are disjoint sets $A_1, \ldots, A_8 \subset S^2$ and $g_1, \ldots, g_8 \in SO(3)$ such that $S^2 = g_1 A_1 \cup \ldots \cup g_4 A_4 = g_5 A_5 \cup \ldots \cup g_8 A_8$

<u>Proof</u>

Let $D$ be as in the Hausdorff paradox (so $D$ is countable). We claim that $\exists \rho \in SO(3)$ such that $\rho^n(x) = \rho^{n'}(x')$ with $x, x' \in D$, $n, n' \in \mathbb{Z}$ $\Rightarrow n = n'$, $x = x'$.

If $\rho^n(x) = \rho^{n'}(x')$, $n \neq n'$, then the fixed axis of $\rho$ lies on the hyperplane bisecting $x, x'$. But the union of countably many planes in $\mathbb{R}^3$ is not all of $\mathbb{R}^3$ (proof via Baire Category).

Consider $B_1 = S^2 \setminus (D \cup \rho D \cup \rho^2 D \cup \ldots)$

$B_2 = \rho(D) \cup \rho^2(D) \cup \ldots$ Then $B_1 \cup B_2 = X = S^2 \setminus D$

but $B_1 \cup \rho^{-1} B_2 = S^2$.

Let $A_1, \ldots, A_4, g_1, \ldots, g_4$ be as in the Hausdorff paradox.

Define $C_{ij} = A_i \cap g_i^{-1} B_j$, $i = 1, \ldots, 4$, $j = 1, 2$.

These 8 sets are disjoint.

$$S^2 = g_1 C_{11} \cup \rho^{-1} g_1 C_{12} \cup g_2 C_{21} \cup \rho^{-1} g_2 C_{22}$$
$$= g_3 C_{31} \cup \rho^{-1} g_3 C_{32} \cup g_4 C_4 \cup \rho^{-1} g_4 C_{42}$$

This is 8 sets, 8 rotations of the type claimed.

<u>Proposition</u>

There a disjoint sets $A_1, \dots, A_8 \subset S^2$ and $g_1, \dots, g_8 \in SO(3)$ such that $S^2 = g_1 A_1 \cup \dots \cup g_4 A_4 = g_5 A_5 \cup \dots \cup g_8 A_8$

<u>~~Propos~~ Proof</u>

As above with $S^2$ replaced by $B^3 \setminus \{\underline{0}\}$. Decompose into spherical shells.

<u>Proposition</u>

There are disjoint sets $A_1, \dots, A_{16} \subset B^3$ and $g_1, \dots, g_{16} \in \mathrm{Isom}^+(\mathbb{R}^3)$, such that $B^3 = g_1 A_1 \cup \dots \cup g_8 A_8 = g_9 A_9 \cup \dots \cup g_{16} A_{16}$

<u>Proof</u>

Let $\Theta$ be an irrational rotation about $(0, 0, \frac{1}{2})$

Define $B_1 = B^3 \setminus \{0, \theta(0), \theta^2(0), \dots\}$ ← distinct points in $B^3$

$B_2 = \{\theta(0), \theta^2(0), \dots\}$

Note that $B_1 \cup B_2 = B^3 \setminus \{0\}$ and $B_1 \cup \theta^{-1}(B_2) = B^3$.

Now repeat what was done before.

<u>Theorem</u>

We can decompose $B^3$ into 17 disjoint sets which may be rearranged to form two copies of $B^3$.

Proofs

disjoint union

Let $X = A_1 \cup \dots \cup A_{16}$. Consider the maps $f : X \to B^3 \sqcup B^3$

$\Psi : B^3 \to B^3 \sqcup \emptyset$

Set $S_1 = f^{-1}(\Psi(S_0))$

$S_2 = f^{-1}(\Psi(S_1))$ etc


Define $S = \bigcup_{i=0}^{\infty} S_i$

Consider $A_i' = A_i \setminus S$ for $i = 1, \dots, 16$, and $A_{17} = S$. This is a decomposition of $B^3$ into 17 pieces and $f(X \setminus S) = (B^3 \sqcup B^3) \setminus \Psi(S)$

Thus $A_1', \dots, A_{16}'$ rearrange to form $(B^3 \sqcup B^3) \setminus \Psi(S)$.

$S$ is a rigid translate of $\Psi(S)$ $\square$

(See Cantor - Schröder - Bernstein, "back and forth")

## The Brunn-Minkowski inequality and Isoperimetric inequality

We will ~~understand~~ assume that we understand the <u>volume</u> of open subsets of $\mathbb{R}^n$ (using only $n=2$). We will write $|V| = \mathrm{vol}(V)$

**Theorem** (Brunn-Minkowski)

Let $A, B \subset \mathbb{R}^n$, open and bounded. Define $A+B$ to be $\{a+b \mid a \in A, b \in B\}$ (easily seen to be open).

Then $|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}$

**Remark**

We have equality if $A, B$ are convex.

**Isoperimetric Inequality**

Suppose $A \subset \mathbb{R}^n$. We define the <u>Minkowski Surface Area</u>, $|\partial A|$ as follows: Let $S \subset \mathbb{R}^n$ be the unit ball, and define $|\partial A| = \lim_{\varepsilon \to 0^+} \frac{|A+\varepsilon S| - |A|}{\varepsilon}$. When $A$ is "nice" this corresponds well with the intuitive notion of surface area.

For notational simplicity (more or less the only reason), set $n=2$.

**Theorem** (Isoperimetric Inequality)

Let $A \subset \mathbb{R}^2$. Then $\frac{|\partial A|}{|A|^{1/2}} \geq \frac{|\partial S|}{|S|^{1/2}}$

(Suitably normalised, nothing has smaller surface area than $S$).

**Proof**

We have $\frac{|A+\varepsilon S| - |A|}{\varepsilon} \geq \frac{(|A|^{1/2} + \varepsilon |S|^{1/2})^2 - |A|}{\varepsilon}$

(Brunn-Minkowski with $B = \varepsilon S$. Note $|B| = \varepsilon^2 |S|$)

Hence, expanding out, $|\partial A| = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(|A+\varepsilon S| - |A|)$

$$\geq \lim_{\varepsilon \to 0} (2|A|^{1/2}|S|^{1/2} + \varepsilon |S|) = 2|A|^{1/2}|S|^{1/2}$$

On the other hand, $|\partial S| = \lim\limits_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\frac{|S+\varepsilon S| - |S|}{\varepsilon}\right)$.

(But $S+\varepsilon S = (1+\varepsilon)S$) $= \lim\limits_{\varepsilon\to 0} \frac{(1+\varepsilon)^2 - 1}{\varepsilon}|S| = 2|S|$

Thus $\frac{|\partial A|}{|A|^{\frac{1}{2}}} \geq 2|S|^{\frac{1}{2}} = \frac{|\partial S|}{|S|^{\frac{1}{2}}}$ $\square$

Proof of Brunn-Minkowski

Idea: Verify the case when $A, B$ are finite unions of boxes, by induction on the number of boxes.

Base Case

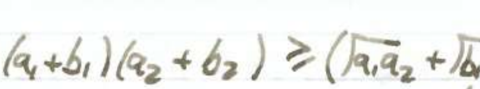
$A$ and $B$ are both boxes. Suppose that $A$ has dimensions $a_1, a_2$, and $B$ has $b_1, b_2$. Then $|A+B| = (a_1+b_1)(a_2+b_2)$
$|A| = a_1a_2$, $|B| = b_1b_2$.

We need to check that $|A+B| = (a_1+b_1)(a_2+b_2) \geq (\sqrt{a_1a_2}+\sqrt{b_1b_2})^2 = (|A|^{\frac{1}{2}}+|B|^{\frac{1}{2}})^2$

Rearranging, this is equivalent to $a_1b_2 + b_1a_2 \geq 2\sqrt{a_1a_2b_1b_2}$
i.e. $(\sqrt{a_1b_2} - \sqrt{a_2b_1})^2 \geq 0$. (Remark: use AM-GM ~~and~~ or other inequalities for the $n$ dimensional case).

Inductive Step

Suppose that $A$ has $\geq 2$ boxes. Find a (horizontal or vertical) line in $\mathbb{R}^2$ that divides $A$ properly,




so that if $A$ has $n$ boxes, ~~then there are fewer than $n$ boxes both below and above the line, i.e.~~ then $A^+$, $A^-$ have strictly fewer boxes than $A$, where

$A^+$ = union of boxes above the line

$A^-$ = union of boxes below the line

Now move $B$ up or down if necessary in such a way that $\frac{|A^+|}{|A|} = \frac{|B^+|}{|B|}$ where $B^+$ is the intersection of $B$ with the points above $L$. This is possible by continuity: if $B$ is moved to $+\infty$ this ratio $\to 1$ and if $B$ is moved to $-\infty$, the ratio $\to 0$, so apply the Intermediate Value Theorem.

Now $|A+B| \geq |A_+ + B_+| + |A_- + B_-|$ (since $A_+ + B_+$, $A_- + B_-$ lie on opposite sides of $L+L$, so are disjoint.

Both $A_+ \cup B_+$ and $A_- \cup B_-$ have fewer boxes than $A \cup B$ since $A_+$, $A_-$ have fewer boxes than $A$.

Applying the induction hypothesis, this is at least

$$(|A_+|^{\frac{1}{2}} + |B_+|^{\frac{1}{2}})^2 + (|A_-|^{\frac{1}{2}} + |B_-|^{\frac{1}{2}})^2$$

$$= |A_+|\left(1 + \sqrt{\frac{|B_+|}{|A_+|}}\right)^2 + |B_+|\left(1 + \sqrt{\frac{|B_-|}{|A_-|}}\right)^2 \qquad (*)$$

However $\frac{|A_+|}{|A|} = \frac{|B_+|}{|B|}$ and so $\frac{|A_-|}{|A|} = 1 - \frac{|A_+|}{|A|} = 1 - \frac{|B_+|}{|B|} = \frac{|B_-|}{|B|}$

and so $\frac{|B_+|}{|A_+|} = \frac{|B_-|}{|A_-|} = \frac{|B|}{|A|}$

Hence $(*)$ simplifies to $(|A_+| + |A_-|)\left(1 + \sqrt{\frac{|B|}{|A|}}\right)^2 = (\sqrt{|A|} + \sqrt{|B|})^2$

QED.

Topics in Analysis

# Besicovitch Sets and the Kakeya Problem

## Kakeya Problem

What is the area of the smallest set $E \subset \mathbb{R}^2$ in which one can rotate a (thin) unit rod through $180°$?

## Answer

For every $\varepsilon > 0$ there is a set $E$ with this property and $|E| < \varepsilon$. We will not quite show this, but will construct a closely related object.

## Theorem (Besicovitch)

There is a compact set $E \subset \mathbb{R}^2$ which contains a unit line segment in every direction but has measure $0$.

## Remark

It is 'easy' to specify a set $E$ with this property, but surprisingly hard to show that it has measure $0$.

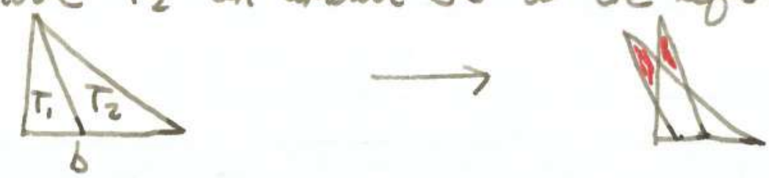
$C = \{\text{base 4 Cantor Set}\} = \{ \sum_{i \geq 1} a_i 4^{-i} : a_i \in \{0, 1\} \}$

$2C = \{ \sum_{i \geq 1} b_i 4^{-i} : b_i \in \{0, 2\} \}$

Join each point on the bottom to every point on the top with a straight line. Since everything in $[0,1]$ can be written as a difference of something in $C$ and something in $2C$, the resulting set has a unit line segment of every angle $\theta$ within $\frac{\pi}{4}$ of vertical. The union of 10 rotations of this set has a line segment in every direction. It is true, but tricky to show, that this set has measure $0$.

Geometric Lemma

Let $T$ be a triangle with base $b$ on the real line and with height 1. Bisect the base, giving two triangles $T_1$ and $T_2$. Move $T_2$ an amount $\delta b$ to the left as shown:




This gives a new triangle with area $(1-\delta)^2|T|$. This is similar to $T$ but the base is shrunk by a factor $1-\delta$. We also have the red "bowtie" shown, which has area $2\delta^2|T|$ (To prove this, dilate so that $T$ is an isosceles right triangle and then use Euclidean geometry).

Lemma

Start with a triangle $T$, contained in an open set $V$. Let $\eta > 0$. Then, for some $k$, we can divide the base of $T$ into $2^k$ parts, which we can then slide around to give a set $E$, which is a union of $2^k$ triangles, with total area less than $\eta$. Furthermore, we can assume that $E \subseteq V$.

Proof

Let $\delta = \eta/10$. Let $k$ be a quantity to be specified later. Suppose WLOG that $|T| = 1$. Divide the base of $T$ into $2^k$ equal parts, $T_1, \dots, T_{2^k}$. Apply the procedure from the Geometric Lemma to the pairs $T_1, T_2$, then $T_3, T_4$, and so on. The total area of the "bowties" is $\delta^2$

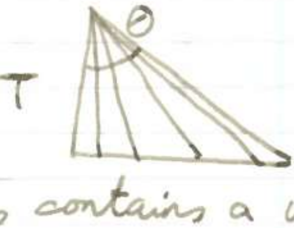
We have new triangles $T_1', T_2', \ldots, T'_{2^{k-1}}$ of total area $\frac{1}{2}(1-\delta)^2$. Now we apply the same procedure to $T_1'$ and $T_2'$, $T_3'$ and $T_4'$, and so on, moving the "bowties" with the triangles. Repeat this until it has been done $k$ times in total.

At the end of the process, we have a single triangle of area $\frac{1}{2}(1-\delta)^{2k}$, and a union of "bowties", with total area at most

$$\delta^2\left(1 + (1-\delta)^2 + (1-\delta)^4 + \ldots + (1-\delta)^{2k}\right) \leq \frac{\delta^2}{(1-\delta)^2} \leq 4\delta$$

if $\delta$ is small. Choosing $k$ large enough, we can make the total area $\leq \eta$. Furthermore, the resulting set has a unit line segment in all the directions that $T$ did.

Finally, we remark on how this can be done while staying in $V$. Since $T$ is compact and $V^c$ closed, some neighbourhood $N_\varepsilon(T)$ also lies in $V$. At the beginning of the construction, divide $T$ into finitely many triangles of base $< \varepsilon$. Then perform the construction as described, noting that nothing was moved by more than $\varepsilon$.




<u>Conclusion of Proof</u>

Begin with a right isosceles triangle $T$. This contains a unit line segment for each angle $\theta \in [0, \frac{\pi}{4}]$. Put $T$ inside some open $V_0$. Now perform the construction, to get a union of triangles $T_1 \subset V_0$, and area $< \frac{1}{10}$. Let $V_1$ be an open set containing $T_1$ and with $\overline{V_1} \subseteq V_0$. Now repeat the construction

on $T_1$, getting a union of triangles $T_2 \subset V_1$, $|T_2| < \frac{1}{100}$.
Continue in this fashion, ensuring that $|V_i| \leq 2|T_i|$, so $|V_i| \to 0$ as $i \to \infty$.

Finally, let $F = \bigcap \overline{V_i}$. We claim that $F$ contains a unit line segment in every direction $\theta \in [0, \frac{\pi}{4}]$. By construction, each $T_i$, and hence $\overline{V_i}$, contains a line segment $X_i + e^{2\pi i\theta}[0,1]$. By sequential compactness, we may pass to a subsequence with $X_i \to X$. Since $F$ is closed, it contains $X + e^{2\pi i\theta}[0,1]$.
Since $F \subset V_i$, $|V_i| \to 0$, $F$ has measure $0$. Finally, take 8 rotated copies of $F$. $\square$

By taking products of planar Besicovitch sets with $d-2$ dimensional planes, we can obtain similar examples in $\mathbb{R}^d$.

<u>Unsolved Problem</u>

Does a Besicovitch set in $\mathbb{R}^d$ have Minkowski dimension $d$?

$$\cancel{\sup_s\{\delta^{-s}|N_\delta(E)}\ d - \sup_s \{ s : \lim_{\delta \to 0} \delta^{d-s} |N_\delta(E)| = 0 \}$$

This is solved when $d=2$. When $d=3$, the best bound is $\frac{5}{2} + 10^{-10}$ (Katz, Łaba, Tao)