Notes on Miscellaneous Mathematical Topics

Thomas Forster

May 12, 2025

Contents

Chapter 1

Stuff to fit in

Dan T sez: add π to he rationals, you get a field that is a finite extension. But is it a finite extension as a ring??

A conversation with Imre and his lads:

Can every planar graph be drawn in the plane in such a way that the length of every edge is rational? Would we prove this by induction?

Show that there is a triangle in the plane all of whose sides are of rational length and that no point in the plane is a rational distance from all three vertices. Try the triangle $(0,0),(0,1),(1/2,\sqrt{3}/2)$.

Imre sez: for any [finite?] graph G there is an n such that every graph that G does not embed into can be drawn on a surface of genus n. (with a small amount of special pleading)

Hermann sez: any rotation is the composition of two reflections. Better still, the angle thru' which you are rotating corresponds to the angle between the two submanifolds you are reflecting in. Think about this in one, two or three dimensions. I suppose in one dimension it is saying that any translation is a product of two reflections. Isn't any isometry a product of two reflections?

Any point on the sphere can be moved to any other point by a single rotation. I can identify a region (of the surface) of the sphere by two points. So we should not be surprised to learn that by two rotations i can move a continent to anywhere else on the globe. 4/xii/24

Always divide by $2\pi i$ just to be on the safe side.

The term "monstrous moonshine" was coined by Conway, who, when told by John McKay in the late 1970s that the coefficient of qq (namely 196884) was precisely one more than the degree of the smallest faithful complex representation of the monster group (namely 196883), replied that this was "moonshine"

(in the sense of being a crazy or foolish idea).[b] Thus, the term not only refers to the monster group M; it also refers to the perceived craziness of the intricate relationship between M and the theory of modular functions.

1.1 An associative metric on the reals?

Can there be a metric on the reals which – tho'rt of as a function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ – is associative? We will write it with 'd', as usual. We can think of d as a function defined on finite multisets of reals. I tho'rt i had proved that it was impossible, but i hadn't. It's a useful question beco's it is so obviously batty.

Evidently $d(x, x, 0) = d(d(x, x), 0) = d(0, 0) = 0$. So $d(d(x, 0), x) = 0$, whence $d(x, 0) = x$ (co's $d(x, y) = 0 \longleftrightarrow x = y$).

Now suppose $d(x, y) = d(x, z)$. Then $d(x, x, y, z) = 0$ and $d(y, z) = 0$ and $y = z$.

1.2 Waiting for buses

I seem to remember a theorem that says that if a bus service runs on average every 10 minutes, and it is a poisson process, then the average expected wait for the next bus is 10' not the 5' that you probably suppose. This seems to fit with experience, but i am trying to reconstruct it. I seem to remember that you know about probability, and now that Ted Harding is no longer with us you are the closest thing i have to a tame probabilist.

Take the open unit interval $(0, 1)$. Sprinkle n blue dots over it. Sprinkle n red dots on it. For each blue dot measure the distance between it and the first red dot to its right. What sort of distribution do these distances have? Mean, standard deviation etc..? Can you shed any light on this..?

Matt replies

Hi Thomas,

One answer is almost definitional – see Grimmett and Stirzaker. If $N(s)$ is the number of buses that have passed by a certain point in time s, with nominal rate 1 every 10 minutes, then by definition of a Poisson process

$$Pr(N(t) - N(s) = 0) = e^{((t-s)/10)}$$

(for $t \geq s$). Supposing you arrive at time s, then

$$Pr(N(t) - N(s) = 0) = Pr(\text{waitingtime} > (t - s))$$

from which we can identify that the waiting time follows an exponential distribution with rate 1/10. This has mean and standard deviation 10.

Note that the waiting time satisfies the Markov property... it is time elapsed from the most recent observation that counts and not the arrival time of the most recent bus. Of course, in practice, buses on the same route "interact" with each other and are not independent...

The way you suggest for constructing a Poisson process of bus arrivals is of course synonymous!

I didn't really follow that...

1.3 Jordan Curve theorem

Apparently it's famously difficult. I wonder if there is a simple proof along the following lines.

Suppose $C \subset \mathbb{R}^2$ is a cts image of the circle. Define an equivalence relation $\sim$ on $\mathbb{R}^2 \setminus C$ by $x \sim y$ iff there is a path from x to y that avoids C. Easy to show that $\sim$ is an equivalence relation. We want to show that it is of index precisely 2. I think it must be of index at least 2 because there must be a compact closed set enclosing C and that must help somehow.

We observe immediately that each $\sim$-equivalence class is a connected open set. Think about the boundaries of these open sets. Their union must be C, which is a closed curve. But each boundary is a closed curve. How can we have two closed curves in the plane one of which is a proper subset of the other?

Or is it perhaps the case that every cts injection from the circle into the plane can be extended to a cts injection from the plane into the plane?

Perhaps the hard part is showing that it cannot be of index > 2.

1.4 Lagrange for Hermann

Let G be a group, H a subgroup of G.

A (right) coset of H is a subset $\{hg : h \in H \wedge g \in G\}$ of G.

To establish Lagrange's theorem we need to establish that the right cosets of H form a partition of G, and that all the pieces of that partition are the same size. This will mean that $|G| = |H|$ times the number of cosets so – at the very least, the size of the subgroup H is a divisor of the size of the group G. (In this culture people talk about the 'order' of the group rather than its size (or cardinality) which is also a bit confusing – since it uses the same word 'order' in 'order of an element' which means the cardinality of the subgroup generated by that element. The word 'order' is – as the compscis say – *overloaded*.)

Let's check that the cosets constitute a partition.

First: every element belongs to a coset.

Let us write 'Hg' for the coset $\{hg : h \in H \wedge g \in G\}$ of G. Strictly speaking, notating a coset in this way gives you extra information about that coset, since it is recording the particular $g \in G$ that gave rise to it. It is worth remembering that a coset might be Hg for more than one $g \in G$.

Anyway, any element g belongs to the coset Hg since H contains the unit.

Now we want to check that any two cosets are either identical or disjoint. The best way to prove this is to establish that any two cosets that overlap are

identical. We do this by explaining how you an "walk around inside" a coset. Suppose a and a' both belong to a coset C. $C = Hg$ for some $g \in G$. So $a = hg$ for some $g \in G$ and some $h \in H$, and $a' = h'g$ for some other $h' \in H$. So $a' = h^{-1}h'a$. Now $h^{-1}h'$ is a member of H, since both h and h' are and H is a subgroup. This says that you can walk around inside a coset, going from any element of that coset to any other element in one hop by multiplying on the left by a member of H. So if two cosets overlap you have a free travelpass.

Finally you need all cosets to be the same size. Why is Hg the same size as Hg'? If you multiply a member of Hg by $g^{-1}g'$ on the right you get a member of Hg'. This mapping is 1-1 beco's group multiplication is *cancellative*: $ab = ac \rightarrow b = c$.

¡¡Do not read this section!!

Now a word about the axiom of choice, which you can skip if you want.

It's all to do with the definition of multiplication of cardinals. If a and b are cardinals, and $a = |A|$ and $b = |B|$ then what is the cardinal $a \cdot b$? We define it to be the cardinal of the cartesian product $A \times B$. So if we want to say that c (which is, say, the cardinal $|C|$ of a set C) $= a \cdot b$ then we have to find sets B and A s.t. C is the same size as – *i.e.*, is in bijection with – $A \times B$. Consider our current situation carefully. . . .

We want to say that $|G| = |H|$ times something. So there has to be a set C s.t. G is the same size as $H\times$ something. So we have to be able to marry up each $g \in G$ with an ordered pair of an $h \in H$ and a . . . a what? The obvious thing would appear to be the coset that contains g. But then we have to know how to recover g from the two things: the coset it belongs to and the $h \in H$. But there is no clear way to do this. If I give you a member h of H and a coset C, which element of G are you supposed to recover from h and C? The point is that each coset can be obtained from H in lots of different ways: every time you multiply H on the right by a member of G you get a coset . . . but there are $|G|$ many members of G – and that is much more than the number of cosets, so each coset will arise in more than one way. If I give you h in H and a coset C, presented as Hg, then that identifies the element hg of G. But for this system of notations to work I have to be thinking of each coset C as Hg for some g depending on C and there are typically lots of g that give rise to any one C. So your choice of g is not obvious.

But couldn't we have chosen to define multiplcation differently? Could we have said that $a \cdot b$ is the cardinality of the union of a set A of pairwise disjoint things each of size b? Notice a contrast between this and the definition we have been using. If $a \cdot b$ is to be the size of a cartesian product $A \times B$ where $|A| = a$ and $b = |B|$, we find that what $a \cdot b$ turns out to be does not depend on which A and B we choose but depends only on their sizes. If we define $a \cdot b$ in the novel way being suggested then it *does* depend on features of A and B other than their size. That is the point of the story about pairs of shoes and pairs of socks. The union of $\aleph_0$ things each of size 2 might be of size $\aleph_0$ (shoes) but it might not (socks). Another way of putting it is that equipollence is a congruence relation for cartesian product but it is not a congruence relation for the other operation . . . which is actually quite hard to describe. Could say quite a bit about this...

But don't worry about that!!

1.4.1 The Factorisation Problem

We are trying to factorise a large number n. Randomly choose a number $x < n$. (It's the *choice* not the *number* that is random – there is no such thing as a random number!) Ascertain (somehow!) the r s.t. $x^r \equiv 1 \pmod{n}$. (This quantity r is customarily called the *order* of r. It's the cardinality of that subgroup of invertible elements mod n that is generated by x. Since people in this business say 'order' when they mean 'cardinality' – or *size* – they end up applying that same word 'order' to the element that generates that group, so they say that the element has 'order' r when they mean that the subgroup that it generates has 'order' (which is to say, *size*) r.)

If r is odd we've got a dud, but if r is even we can reletter 'r' as '$2r$' and reflect that what we learn from $x^{2r} - 1 \equiv 1 \pmod{n}$ is that n divides $(x^r - 1)(x^r + 1)$. So every prime factor of n divides precisely one of $x^r - 1$ and $x^r + 1$. Naturally we run Euclid's algorithm on $(n, x^r - 1)$ and on $(n, x^r + 1)$. If n is prime then one of these processes will return n and the other will return 1, which is no use. If n is composite we just cross our fingers and hope that not all its factors fall into the one bucket, beco's if they fall into more than one bucket then the two calls to Euclid will both return nontrivial factors of n. Of course it might be that n divides one of $x^r - 1$ and $x^r + 1$ even tho' it is composite, and that's just our bad luck.

So the idea is that we pick numbers $x_1\, x_2 \ldots$ less than n at random, and run Euclid on each (x_i, n). If we get a factor we are happy. If we don't, we calculate the "order" of x_i, and if the order is even we do the above trick. The more often we do this, the better our chance of detecting that n is composite – if it is!

Of course if the order of x_i is not just even but has a significant power of 2 as a factor we get a second – perhaps even a third – bite of the cherry for nothing. Suppose x is of order $8r$. Then we have $x^{8r} - 1$ is divisible by n. But we can factorise $x^{8r} - 1$ further than we can factorise $x^{2r} - 1$. We can factorise it into

$$(x^{4r} + 1)(x^{2r} + 1)(x^r + 1)(x^r - 1)$$

giving us *four* numbers that we can pop into the Euclid mincer with n, not just two – giving us an extra chance of finding a nontrivial factor if there is one. After all, if n has more than one factor then those different factors might divide into different numbers from the set $\{x^{4r} + 1,\, x^{2r} + 1,\, x^r + 1,\, x^r - 1\}$.

1.4.2 Why do we include the inverse operation in the definition of a group?

It's all to do with the definition of a substructure of a structure.

A structure is a set with knobs on. A Group is a set with three knobs: a designated element (the unit) a two-place operation (multiplication) and one one-place operation (inverse). A ring is a set with two further knobs, co's it has

a multiplicative unit and an second multiplication operation (the first multiplication is now called *addition*) but we don't add an extra knob for multiplicative inverse.

A substructure of a structure is a subset of the structure equipped with the restriction of the knobs. Thus a substructure of a group G is a subset of G equipped with the restriction of the knobs. So it has the unit element, and multiplication and inverse. And – and this is the crucial part of the definition of substructure – it must be closed under the restriction of those operations; the substructure must be closed under multiplication and inverse. Now! If i am a substructure of a group containing the unit element and closed under multiplication and inverse then i am a group!

In general the idea is that one should define groups, rings, wombats etc in such a way that a substructure of a group, ring, wombat etc is another group, ring, wombat etc. So if (when setting up group theory for example) you want every element to have an inverse, then you'd better write the inverse operation into the spec. It is possible to axiomatise group theory in the smaller language without the function symbol for inverse, but if you do it is no longer true that a substructure of a group is a group. It will be a semigroup with a unit and cancellation (don't ask) but cannot be relied upon to be a group. I should be able to think of a cute illustration, but the only natural example that comes to mind (perhaps beco's i am thinking about computability) is the group of computable permutations of $\mathbb{N}$; the collection of primitive recursive permutations of $\mathbb{N}$ is a substructure of it closed under composition (and contains the identity element) but – incredibly – is not closed under inverse[1]. So it's not a subgroup. It's a subsemigroup but it's not a subgroup. If you set up group theory without the inverse function then this subsemigroup is a substructure but it's still not a subgroup. So you would have a substructure of a group that is not a group.

> It's not obvious that the inverse of a primitive recursive permutation is primitive recursive;
>
> It's not obvious that the inverse of a setlike permutation is setlike'
>
> It's not obvious that if σ and τ are skew-conjugate so are σ^{-1} and τ^{-1}.

Are these facts related?

A conversation with John Howe 4/xi/17

He says;

OK, as you say, the class of integral domains is the class of substructures of fields. What about the class of substructures of groups-without the inverse operation? It's the class of left-and-right cancellative semigroups with $\mathbb{1}$: $(\forall abc)(ab = ac \to b = c)$ and $(\forall abc)(ba = ca \to b = c)$. Now any integral domain has a canonical extension to a field. How about cancellative thingies?

[1] Don't ask.

Take your cancellative thingie, take a name for each element, and consider the group generated by the names with the obvious equations given by the diagram of the cancellative thingie. Clearly the cancellative thingie embeds in the group. But is the group the minimal group that does what we want?

1.5 NZMRI Napier jan 2021

A problem is complete for a class Γ if every problem in the class is **many-one** reducible to it. (not Turing-reducible). $0^{(n)}$ is Σ^0_n-complete.

Kleene-Spector: Set of wellorders is Π^1_1-complete.

'completely decomposable abelian group' when restricted to countable structures is Σ_7.

Klein's criterion aka the Ping-Pong Lemma.

Set of automatic structures is Σ^1_1-complete.

1.5.1 Gabriel Verret on vertex-transitive graphs

semiregular: at most one g sending x to y. (property of the group action not of the group)

transitive groups are big, semiregular grps are small.

regular is transitive and semiregular. The rotation group on a polygon is regular.

$\tilde{g}$ is mult on the R by g. $\{\tilde{g} : g \in G\}$ is a regular group (acting on G)

A grp is vertex-transitive iff $\mathrm{Aut}(G)$ is transitive on the vertices

Petersen graph set of doubletons from $\{1, 2, 3, 4, 5\}$ edge between doubletons that are disjoint.

Cayley graph not Kayleigh graph! There should be a joke there.

$\mathrm{Cay}(G, S)$ is connected iff $G =< S >$

All Cayley graphs are vertex-transitive

Sabidussi 1958

If Γ is a graph and G is a regular subgroup of $\mathrm{Aut}(\Gamma)$ then Γ is $\mathrm{cay}(G, S)$ for some S.

True in the infinite case too.

Godsil 1978

1.5.2 Codes and Geometry

Fundamental problem in coding theory is that you want to optimise length, dimension and distance simultaneously! perfect codes are codes meeting the Hamming bounds

Fanocode. 4 bits plus 3 for checksums

Does the family of Boffa atoms form a geometry?

for any two bas there is a unique one containing both? For any two there is a unique one belonging to both?

every line houses the same number of points; all points lie on the same number of lines.

Barwise solution lemma!

1.5.3 Marie Graff: Imaging

Apple seismology can measure firmness Physics Today 2017

Should be able to describe inverse problems in terms of translations.

A typical imaging problem is solved by a series of measurements followed by inferences.

1.6 Bruce explains Basic Electricity to me, 10 viii 2020

If you move a charge through a voltage difference of v volts then you are doing some work, and the energy that the charge thereby acquires is v times the charge. Charge is measured in coulombs.

(Sounds like some equivocation here. 'A charge' – the thing you are moving – sounds like a count noun. but the stuff that is measured in Coulombs is a quantity. Perhaps the first sense is an object that is decorated with the second.)

Strength of an electric field is volts divided by distance. Bruce doesn't know a CGS unit for this. (same as newtons per coulomb)

Power is energy per second, MKS (not CGS any more smackie handie) unit is watts.

Power = volts $\times$ (charge per second). Compare getting energy from falling water. Voltage corresponds to height difference, as i woz told when i woz little. Charge corresponds to the weight of water falling (NOT the mass, co's it's force = mass $\times$ g (which in general is $G \cdot (M_m/r^2)$). Less energy on a less massive planet.)

Current = charge/second = amps.

Should connect this somehow with the fact that the energy extracted by a windmill increases with the *cube* of the wind speed.

volts $\times$ current = power (watts); joule/second = watt.

joules = coulombs $\times$ volts. Written out in primitive notation time does not appear. Same parameter as calories. Calories heat stuff through temperature difference. One calorie heats one gram through one degree centigrade. Nontrivial discovery that joules and calories are the same parameter.

Primitive parameters: charge, time, length, mass.

one watt for one second gives you one joule.

Human body runs on something like 60 watts.

Now i need to understand why mv and mv^2. Since $E = mc^2$ the thing that is mv^2 is presumably *energy*. So what is mv?

The Chinese remainder theorem says that if $n = p \cdot q$ then

$$(\mathbb{Z}/n\mathbb{Z})^* \simeq (\mathbb{Z}/p\mathbb{Z})^* \times (\mathbb{Z}/q\mathbb{Z})^*$$

where the * means the multiplicative group of nonzero elements.

$$x \bmod n \longleftrightarrow \langle x \bmod p, x \bmod q \rangle$$

Robert Honig says: if a field is not of finite characteristic then it's quite correct to say it's of characteristic 0 beco's 0 is the cardinality of the smallest multiset of 1 that sums to 0. That doesn't really make much sense. If there are no such multisets what is the cardinality of the smallest? One could try saying that by "cardinality of the smallest multiset of 1 that sums to 0" what one means is the sum of all the numbers n that are $\leq$ all such cardinals. Nice try. It really would be better to say that fields that are not of finite characteristic have undefined characteristic. But we're never going to win that one.

[TMS] Dr Hamza Fawzi (DAMTP)17/ii/19 - Sum-of-squares proofs

To show $A < B$ express $B - A$ as a sum of squares. e.g. Cauchy-Schwarz

$$(\Sigma a_i^2)(\Sigma b_i^2) - \Sigma a_i b_i = (\frac{1}{2}) \cdot \Sigma (a_i b_j - a_j b_i)^2$$

Is every everywhere +ve poly a sum of squares?
No, Hilbert 1888. Motzkin poly:

$$x^2y^4 + x^4y^2 + 1 - 3x^2y^2$$

Hilbert's 17th problem: is every +ve poly a sum of squares *of rational functions*?
A: yes. (Artin 1927)
Note: $(\forall x \in \mathbb{R})(ax^2 + bx + c \geq 0)$ iff $a \geq 0$ and $(b^2 - 4ac) < 0$

Completeness of Cauchy Reals

Recall that a function $f : \mathbb{N} \to \mathbb{Q}$ is **Cauchy** iff

$$(\forall \epsilon \in \mathbb{R}^+)(\exists n \in \mathbb{N})(\forall m_1 m_2 > n)(|f(m_1) - f(m_2)| < \epsilon)$$

We want to say that two Cauchy sequences are equivalent iff they "converge to the same thing". Then we define a Cauchy real to be an equivalence class of Cauchy sequences. Care is required because it's far from obvious that there is a definable choice function on the set of Cauchy reals. (I'm not saying that there isn't!)

Then the challenge is to define the arithmétic operations and the ordering on the Cauchy reals, and to prove that the result is a complete ordered field. And to do all that without using the axiom of choice!

Probably useful that any cofinal subsequence of a Cauchy sequence is Cauchy. That's true (isn't it...?! Prove it)

A cute fact: two Cauchy sequences are equivalent iff their interleaving is also Cauchy.

Any Cauchy sequence is equivalent to all its cofinal subsequences.

An equivalence class is a $\subseteq$-maximal class of Cauchy sequences closed under interleaving.

So we *define* two Cauchy sequences to be equivalent iff their interleaving is Cauchy. A bit of work to show that this relation is transitive.

Completeness of Cauchy reals needs AC? I hope to show that it doesn't. Completeness of Dedekind reals doesn't; at least i bloody well hope not. What we have to do is: given a Dedekind cut, design a Cauchy sequence that converges to the hole. OK. We have a Dedekind cut, a partition $\{\texttt{left}, \texttt{right}\}$ of $\mathbb{Q}$ where

$$(\forall x \in \texttt{left})(\forall y \in \texttt{right})(x < y).$$

We want a Cauchy sequence $s : \mathbb{N} \to \mathbb{Q}$ that converges to the hole o. Let $s(0)$ be some random member of `left`. Thereafter set $s(n)$ to be largest rational in `left` whose denominator is n (above $s(n)$, presumably). Then $o - s(n) \leq 1/n$.

Also, as Randall points out to me (june 2018) one can do it by reasoning about the Cauchy reals themselves, and without using AC. Suppose you have a Cauchy sequence of (Cauchy) reals; I can't exactly remember his way of doing it, but what follows is my reconstruction.

What is it for a sequence of (Cauchy) equivalence classes to be itself Cauchy? This could be quite tricky! It might be an idea to have a definable choice function on the family of equivalence classes. Each equivalence class E is a set of functions $\mathbb{N} \to \mathbb{Q}$. We can start by cutting down to functions that are monotone increasing. Fix a Cauchy real E (thought of as an equivalence class of functions $\mathbb{N} \to \mathbb{Q}$). Our canonical representative r will be defined by a process rather like that in the minimal bad sequence lemma. Canonically worder the rationals somehow. Set $r(0)$ to be the first rational that is $s(0)$ for some $s \in E$. $r(1)$ be the first rational that is $s(1)$ for some $s \in E$ s.t. $s(0) = r(0)$. Thereafter $r(n+1)$ is the first rational that is $s(n+1)$ for some $s \in E$ s.t. $s(k) = r(k)$ for $k \leq n$. Notice that r is an increasing sequence. We need to check that it is Cauchy and in E. I know it looks pretty obvious but it might be an idea to check. Isn't every E closed in the product topology?

Then, given a sequence $\langle E_n : n \in \mathbb{N} \rangle$ of equivalence classes, we extract from each a representative r_n as above and diagonalise through them to obtain r_∞ defined so that $r_\infty(n) = r_n(n)$.

1.7 Vector Spaces

A Vector space over a field F is an abelian group equipped with a collection of endomorphisms indexed by the field. This indexation is always presented as a binary function: field $\times$ space $\to$ space but i think there is much merit in thinking of it as its curried version: a function from F to a set of homomorphisms.

Does it matter which field element is paired with which homomorphism? The insistence that scalar multiplication should distribute over vector addition says merely that each scalar points to a homomorphism.

What do we want to say about the collection of endomorphisms? I copied this off the board:

$$(-1) \cdot v = -v \tag{1}$$

-1 is the additive inverse of $\mathbb{1}$, the multiplicative unit of the field F; $-v$ is the additive inverse of v in the abelian group. This is quite informative. This formula (1) says that the homomorphism corresponding to the additive inverse of $\mathbb{1}$ is a homomorphism that sends each element of the group to its inverse. But this (perfectly respectable) function is a homomorphism only if the group is abelian. That's why we stipulate that the group should be abelian.

We get a lot of mileage out of our determination to think of vector spaces as algebras. What is a subspace going to be? Well it's obviously a subgroup, but which homomorphisms is it to be equipped with? Well, obviously the restrictions of the homomorphisms... *all* the homomorphisms. This immediately gives us that an arbitrary intersection of subspaces is another subspace. Is there an addtion operation on subspaces? What is $U + V$? It's not going to be $U \cup V$ for the simple reason that that might not be a group. It has to be the subgroup generated by $U \cup V$. The usual definition is $\{u + v : u \in U \wedge v \in V\}$, and even that works only as long as the group is abelian[2] OK, so that's binary sups; what about arbitrary sups? Well, once we reflect that an arbitrary intersection of subspaces is another subspace we are home and hosed. But that's a top-down definition – is there a nice bottom-up definition? Yes, but we need to be careful.

Suppose we have a family $\{U_i : i \in I\}$ of subspaces. What is the sup? It's got to be a subspace, so it's at the least a group. Our group operations are finitary, so we don't have a good notion of infinite sums of elements of the group. However we do have a good notion of sum of an infinite (multi)set of group elements [as long as] all but finitely many of which are the unit of the group, and that's what we use. It would probably be a good idea to write out a proof that this gives us the same answer as the "top-down" definition. And it underlines what a radical departure it is to allow infinite sums.

Mazes

I can't remember who showed me this... it might have been Richard Kaye. Mazes. You make a maze by putting up walls in a chessboard. I think the idea is to get from the bottom left square to the top right square but i suspect the exact details don't matter.

Some mazes you can get through and some you can't. A **sequence of moves** is a word over the alphabet {`north`, `south`, `east`, `west`}. A sequence is allowed to be infinite. Is there a sequence of moves that solves all solvable

[2]Observe that the cardinality of the subgroup generated by $U \cup V$ is bounded by $|U|$ and $|V|$. In fact it is of course $|U| \cdot |V|$ – beco's the group is abelian. Do we get (more relaxed) bounds if the group is nilpotent?

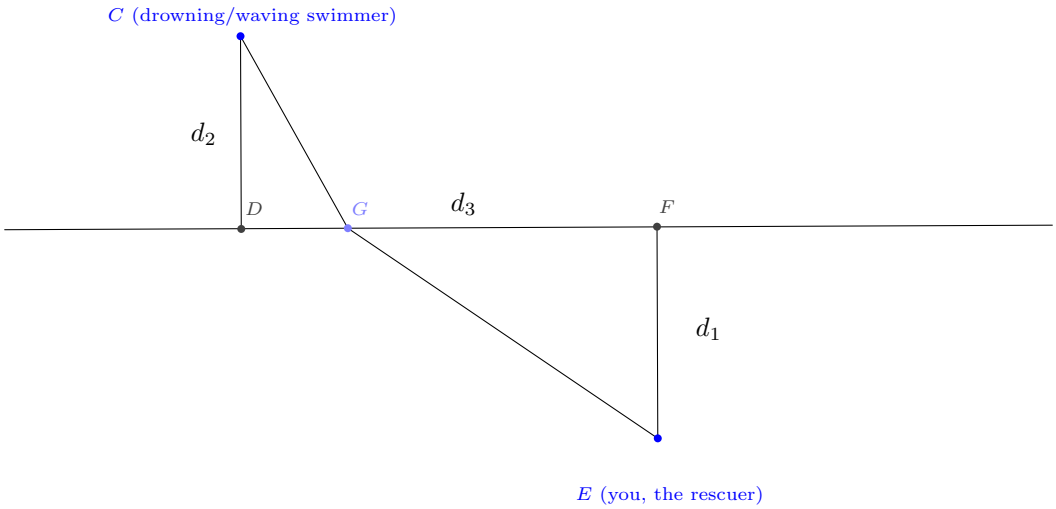
mazes? Outrageously yes there is. Solve the first maze (if your move is to go `north` and that would take you through a wall you stay where you are). Then use the same sequence of moves in the second maze. Continue till you solve the second maze. Then do the third, and so on. [This sounds a bit elliptical to me but i think the key observation is that if a maze is solvable it is solvable from any accessible point. Is this a point about *confluence*...?]

Snell's Law (or rather Thomas Hariot's Law)

A conversation with (the recently departed) John Barrow over lunch back on thursday 23/iii/17. He was talking about Maupertuis and possible worlds.

You are on the beach, and there is someone out at sea that you have to reach to rescue. Speed is of the essence of course, so what path do you follow? The point is that travel across the beach is faster than travel through the water. I'm pretty sure you can solve this just using A-level maths, so i just might be able to do it.

Suppose you are a distance d_1 from the shoreline, the drowning (not waving) swimmer is a distance d_2 from the shoreline, and the distance between the feet of the two perpendiculars is d_3. Your speed on the beach is v_1 and in the water is v_2.



Let α be the angle FEG. Then the time taken by the rescuer is $\frac{d_1 \cdot sec(\alpha)}{v_1}$ (which is the time taken to get from E to G) plus the time taken to traverse the hypotenuse CG. The length of CG is $\sqrt{(d_3 - d_1 \cdot sin(\alpha))^2 + (d_2)^2}$ so the time taken to traverse it is $\frac{\sqrt{(d_3 - d_1 \cdot sin(\alpha))^2 + (d_2)^2}}{v_2}$.

Let's simplify the stuff under the square root sign
$(d_3)^2 - 2 \cdot d_3 \cdot d_1 \cdot sin(\alpha) + (d_1)^2 \cdot sin^2(\alpha) + (d_2)^2$
It all looks a bit messy.

But i am now (april 2021) moved to think about this in a different way, beco's of a conversation i was having earlier today with an interesting but probably crazy physicist i know.

The challenge he has given me (or i have given myself) is find a metric for the bit of shore and sea such that according to that metric, the trajectory i choose to reach the swimmer from my place on the shore is a geodesic. A *straight line*. Can you point me to anything to read on this?

Binary relations on a set of size n with $n \in \mathbb{N}$

Two banal observations:

(i) the number of partial orders is odd. The converse R^{-1} of a partial order R is a partial order and $R \neq R^{-1}$ unless $R = \mathbb{1}$, the identity relation.

(ii) Each isomorphism class is of size $!n$ at most (and that only once, beco's it happens only for total orders) so there are at least $2^{n^2}/!n$ isomorphism classes.

A nice titbit from Imre

Exhibit a bijection $f : \mathbb{N} \longleftrightarrow \mathbb{N}$ s.t. whenever $X \subseteq \mathbb{N}$ is an A.P. of length 3 then $f``X$ is not an A.P. of length 3.

James' talk to the kiddies on Congruent numbers, elliptic curves and L-functions

During one of my visits to CMU.

See Tunnel: Inventiones Math. **72** (1985) fasc. 2 pp 323-334.

Koblitz: Intro to elliptic curves.

A **Congruent Number** is the area of a right-angled triangle with sides of rational length.

The area of the triangle with sides $x^2 - y^2$, $2xy$, $x^2 + y^2$ is $xy(x^2 - y^2)$. So: if $p^2 | xy(x^2 - y^2)$, then the triangle with sides $(x^2 - y^2)/p$, $2xy/p$, $(x^2 + y^2)/p$ is a triangle with area $xy(x^2 - y^2)/p^2$. So $xy(x^2 - y^2)/p^2$ is congruent as long as $p^2 | xy(x^2 - y^2)$.

P is the point $(-1, 0)$. Drive a line through it of slope s and see where it meets the unit circle (of radius 1 centred at the origin). It meets it at the point $Q = ((1 - s^2)/(1 + s^2), 2s/(1 + s^2))$. e.g. $s = 1/5$ gives $(12/13, 5/13)$.

So if n_0 is congruent, so is $n_0 \cdot m^2$ for all m. So it's sufficient to identify *quadratfrei* congruent numbers.

Clearly a correspondence between triples $\langle X, Y, Z \rangle$ with $X^2 + Y^2 = Z^2$, $XY/2 = n$ and numbers x such that $x - n$, x and $x + n$ are all perfect squares.

Identify squarefree congruent numbers by looking at

$$y^2 = x^3 - n^2 \cdot x$$

... beco's this is $y^3 = x(x-n)(x+n)$

So a congruent number corresponds to a rational point on this curve. Conversely? Not exactly.

If P and Q are rational points on the curve then $P + Q$ is the reflection in the x axis of the point where the line PQ meets the curve. Hard to show that it's associative but it is! The point at infinity is the identity element. (What happens if P and Q have the same x-coordinate?? Does that make them inverses?)

1.7.1 Quantum Computing

This is James again

Start with Hilbert spaces. A Hilbert space is a vector space over the complexes, typically of infinite dimension, equipped with an inner product operation written (x, y) that is linear in the left argument and... well, best write it down!

$(\alpha \cdot x, y) = \alpha \cdot (x, y)$

$(x + y, z) = (x, y) + (x, z)$ (I'm guessing that this is what he meant)

but "conjugate linear" on the right...

$(x, \alpha \cdot y) = (x, y) \cdot \alpha$ where one of the two things on the right should be replaced by its conjugate, but i'm not sure which.

I think it also has the effect that $(y, x) =$ conjugate of (x, y).

We also have a norm: $|x| = \sqrt{(}x, x)$ which is zero iff x is. This also gives a metric: $d(x, y) = |x - y|$.

Finally we insist that the topology arising from the metric should be a complete seperable metric space.

A natural example is the space called L^2 (why is it called this?) This is the set of ω-sequences $\langle x_i : i < \omega \rangle$ of complexes such that the sum of $|x_i|^2$ is defined. (Have i written this last formula down properly? Do i not mean $(x_i)^2$?) Then we take (x, y) to be

$$\sum_{i<\omega} x_i \cdot y_i$$

and this must converge for standard reasons.

1.8 What the **** is a Random Variable??

Ted Harding is attempting to explain random variables to me..

We start with a set S on which there is a measure, which is a partial function from $\mathcal{P}(S)$ to $[0, 1]$. The letter 'S' is intended to connote Sample space. The measure is additive in the obvious way, sends $\emptyset$ to 0 and S to 1, etc. Then a random variable is a function $X : S \to \mathbb{R}$ s.t., for any $\alpha \in \mathbb{R}$, X^{-1}"$\{\alpha\}$ is a

measurable subset of S. And the measure of that preimage is the probability of that subset of reals, in the sense that it is the probability that the random variable X takes value α in that set. This probability might be 0, for example if X is a continuous random variable, but if X is a discrete random variable. This is why the measure of a singleton of a member of S might be nonzero. There are lots of different Ss! For example, in Ted's bookshop example, S is the set of volumes sitting on the shelf.

Given a random variable X and a real α, consider $\{s \in S : X(s) \leq \alpha\}$.

This is a measurable subset of S and has a measure, and this real is to be thought of (is defined to be) the probability $P(X \leq \alpha)$. This defines a function $\mathbb{R} \to \mathbb{R}$ which is monotone nondecreasing, unbounded below 1 approaches 0 as X.. this is the cumulative distribution function of X, typically written $S_X(\alpha)$. This S_X determines (characterises?) the probability distribution of X. $P(\alpha_0 < X \leq \alpha_1) = S_X(\alpha_1) - S_X(\alpha_0)$.

Thus random variables are more-or-less forced to take values in $\mathbb{R}$, or at the very least in things that can be coerced into being reals. In Ted's bookshop example he has to coerce the natural number values of authors, number of pages into being reals. The colour of the book doesn't have order structure so *that* random variable at least doesn't have a cumulative distribution function.

Another of Ted's examples is the random number generator, the dekatron. We use this thing by sampling the number (its *state* really, which is a natural number $< 10^4$) every now and then, on a timescale longer than its cycle time (which is $1''$). We then take the remainder mod one second. Something to think very hard about: if this process is to be thought of as a *random variable* then what is S? S must be the set of consecutive-pairs of samplings. So S is a rather open-ended thing, more like a stream than a set.

1.9 Homework for TWK

Proving that the two definitions of cts fn $\mathbb{R} \to \mathbb{R}$ are equivalent, the $\epsilon - \delta$ definition and the definition involving convergent sequences.

A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point $p \in \mathbb{R}$ if given $\epsilon > 0$ there exists $\delta > 0$ such that if $|p - x| < \delta$ then $|f(p) - f(x)| < \epsilon$.

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be continuous at a point $p \in \mathbb{R}$ if whenever $\langle a_i : i \in \mathbb{N} \rangle$ is a real sequence converging to p, the sequence $\langle f(a_i) : i \in \mathbb{N} \rangle$ converges to $f(p)$.

TWK says

"To a delicately brought up logician the two propositions look entirely different. To the hoi poloi of analysts they look the same.

I think we agree that the first proposition implies the second.

Now suppose the first proposition is false. Then there exists an $\epsilon > 0$ such that, for any $\delta > 0$ we can find an x with $|x - p| < \delta$ and $|f(x) - f(p)| > \epsilon$. In particular we can choose a_n [logician murmers using countable choice] with $|a_n - p| < 1/n$ and $|f(a_n) - f(p)| > \epsilon$.

I think 'logicians have shown' that the use of countable choice is unavoidable."

– TWK

1.9.1 Uniform Continuity

On Sat, 8 Aug 2009, Thomas Forster wrote to TWK

I have got this right, haven't i..

$f : \mathbb{R} \to \mathbb{R}$ is continuous in the interval I iff

$$(\forall x \in I)(\forall \epsilon > 0)(\exists \delta > 0)(|x - y| < \epsilon \to |f(x) - f(y)| < \delta)$$

(*id est*, the value of δ depends on both x and ϵ) and

$f : \mathbb{R} \to \mathbb{R}$ is **uniformly** continuous in the interval I iff

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall x \in I)(|x - y| < \epsilon \to |f(x) - f(y)| < \delta)$$

id est, the value of δ depends only on ϵ.

Am i right?

TWK replies:

Yes perfectly correct. It is a deep theorem that if I is bounded and closed then f continuous implies f uniformly continuous.

If you consider the open interval $(0, 1)$ then $x \mapsto 1/x$ is continuous but not uniformly continuous.

$x \mapsto \sin 1/x$ is bounded and continuous but not uniformly continuous.

This definition [uniform continuity] is TWK's example of something that all mathematicians get, but that people with a syntax module get *instantly* – even if they aren't strong mathematicians.

It's probably worth establishing when this definition became crystal-clear. It would be a nice detail in the story that needs to be told about the gradual intrusion of syntax into mathematics. Extension element problem.

But it's not *really* about syntax, it's about dependency. At least it's possible to tell it in such a way that syntax plays no part.

But then perhaps logic is all about dependency relations. That's why constructive logic is so important, beco's it is even more sensitive to these issues. Think about

$$(\forall x \exists y) R(x, y) \to (\forall x \exists y)(R(x, y) \to (\forall x' \exists y')(R(x', y') \wedge (x = x' \to y = y'))).$$

Indeed if one were to make up a line of patter for constructive logic it would be these issues that one would raise, not the crap about excluded middle.

1.10 Congruence relations for Exponentiation

I am fond of pointing out that congruence-mod-p is a congruence relation for addition and multiplication but not for exponentiation (unless $p = 2$ – which i've only just noticed – and of which more below): $2 \equiv 7 \pmod 5$ but 2^7 and 2^2 are not congruent mod 5.

Is there, in fact, a sensible equivalence relation on $\mathbb{N}$ that is a congruence relation for all three operations? Well: equality is, for a start. But of course that's not what we meant. Any refinement of a congruence relation for `op` is also a congruence relation for `op`. What about meets? Obviously not. However the set of congruence relations for any given operation is closed under unions of chains. So there will be maximal relations of this kind, and we can find them without using AC beco's $\mathbb{N}$ is wordered. Indeed this means that there will be definable such relations.

So here's a challenge: is there a *nice* definable $\subseteq$-maximal equivalence relation on $\mathbb{N}$ which is a congruence relation for addition, multiplication and exponentiation? (Other than congruence-mod-2, that is). (Follow-up question: why do i not already know the answer to this question??)

The property of being a congruence relation for all three is in fact a universal Horn property, since it is the conjunction of the three clauses:

$$(\forall xx'yy')(\langle x, x'\rangle \in R \wedge \langle y, y'\rangle \in R \to \langle x + y, x' + y'\rangle \in R)$$

$$(\forall xx'yy')(\langle x, x'\rangle \in R \wedge \langle y, y'\rangle \in R \to \langle x \cdot y, x' \cdot y'\rangle \in R)$$

$$(\forall xx'yy')(\langle x, x'\rangle \in R \wedge \langle y, y'\rangle \in R \to \langle x^y, x'^{y'}\rangle \in R)$$

(on top of the axioms for an equivalence relation of course).

The universal relation fits the bill, and these conditions are clearly intersection-closed. So it gives rise to a notion of closure. I think any such relation that contains a pair $\langle a, b\rangle$ will extend congruence-mod-$|a - b|$.

I was making the point that congruence-mod-m is a congruence relation for $+$ and $\times$ but not for exponentiation. So if you want to know $x + y$ and $x \times y$ only up to remainder mod m all you need is x mod m and y mod m. You don't need the whole of x and y. Now think about x^y mod m. My student Herbie Bowden says that if you want to know what x^y is mod m you may need the whole of y, but all you need to know about x is its residue mod m. He's right!

Any number congruent to x mod m is $x+mn$ for some $n \in \mathbb{N}$. $(x+nm)^y - x^y$ is divisible by m as follows. Expand $(x+nm)^y$ using the binomial theorem. The only term that does not have m as a factor is the leading term, x^y, which we subtract. So x^y and $(x + nm)^y$ are congruent mod m.

So congruence-mod-m comes closer to being a congruence relation for exponentiation than i had properly appreciated.

This is probably related to a fact you will learn next term (it should really be an exercise) that the set of binary representations of *multiples* of 3 is a regular language but the set of binary representations of *powers* of 3 is not.

Chapter 2

First Year Analysis

Multisets; associativity, commutativity, idempotence. Renormalisation. Absolute and conditional convergence. Conditional convergence is a property of wellorderings up to finite rearrangement. Addition of infinite sums: pointwise addition? Interleaving? What if you interleave a sequence with the stream of 0s? Does the infinite term contain variables or not?

A section on extending semantics. Include a discussion of why infinitary XOR cannot be sanitised

Prologue on datatypes: sets, streams, multisets. 'stream' is a bit of Compsci-speak. An object is of datatype α-`stream` iff when you ask it nicely it gives you a thing of type α and then goes back to being a thing of datatype α-`stream`. Explain $L_{\omega_1,\omega}$.

Analysis, Differential calculus, call it what you will, presents challenges to our understanding by making us think about continuous functions. What is continuity? However there is also the challenge – which to a certain extent can be separated from that challenge – of understanding how the operations that work on one or two arguments – such as $+$ and $\times$ and which can be extended to work on finite bundles of arguments – can be extended further to work on infinite bundles of arguments ... sometimes!

I find i have slipped from one extreme position to another. Once upon a time i was young, and doing Infinite series (Oliver and Boyd minipapperbok) when i was a philosophy u/g. I couldn't understand what all the fuss was about. It all seemed obvious. Now i am an elderly logician, and i *absolutely* see what all the fuss was about and why it is not obvious *at all*. In fact my mistrust of continuous mathematicians is now so deeply ingrained that it wouldn't surprise me to find that they still haven't done it properly even now. I am going to do it myself – properly, of course – and see whether we get the same result. Partly i am writing this up for Mansur Boase, who asked me a relevant question, and also for Bruce McKinney.

Just as one has to emphasise that the fundamental challenge is not to ascertain the values of (for example) infinite sums but is instead to see what possible answers these things might have, so one has to make in a

And why was there a copy of Hardy: *Divergent Series* in the school library at my boarding school – Marlborough?

Perhaps one can say something about the way the factorial function can be extended to arbitrary reals by means of the Γ function.

more general context a story about how something that is ungrammatical (but which our FTPM enjoins us to accept) could be given a meaning.

2.1 A Tower of xs

I want to start off with a beautiful puzzle shown me by Martin Richards.

Suppose $x^{x^{x^{x^{x^{\cdots}}}}} = 2$; solve for x.

There is periodic structure we can exploit.

The exponent on the LHS is $x^{x^{x^{x^{x^{\cdots}}}}}$ which we are told is 2, so $x^2 = 2$ and $x = \sqrt{2}$. That was easy.

Now consider

$x^{x^{x^{x^{x^{\cdots}}}}} = 4$; solve for x.

The problem with this is that this second equation gives $x^4 = 4$ and thence $x = \sqrt{2}$ again. They can't both be right!

Of course the answer is that the reasoning that led us to conclude that $x = \sqrt{2}$ in the first place doesn't prove that that is the answer. All we have done is show that **if** there is a solution it must be $\sqrt{2}$. We haven't shown that there **is** a solution. In fact it is a simple matter to show by induction that the approximants to the LHS, which we generate as follows...

$$a_0 =: \sqrt{2}; \quad a_{n+1} =: \sqrt{2}^{a_n}$$

...are all less that 2. We do this as follows

$$a_{n+1} = \sqrt{2}^{a_n} < \sqrt{2}^2 = 2$$

where the middle inequality follows by induction hypothesis. So the sequence has a limit which is ≤ 2. So 2 is indeed a solution as alleged.

Let $F(x) =_{df} x^{x^{x^{x^{x^{\cdots}}}}}$

We have $x^{F(x)} = F(x)$. The inverse to this function is the function $x \mapsto x^{1/x}$, obtainable as follows.

$$x^y = y$$

$$y \cdot \mathtt{log}(x) = \mathtt{log}(y)$$

$$\mathtt{log}(x) = (1/y)\mathtt{log}(y)$$

$$x = y^{1/y}$$

This is much easier to understand. For example we can differentiate it. It is the same as $e^{(\mathtt{log}\, x)/x}$ whose derivative is of course $e^{(\mathtt{log}\ x)/x} \cdot (1/x^2 - (\mathtt{log}\, x)/x^2)$. This is zero when $x = e$, and this is clearly a maximum. The fact that the derivative is zero there of course means that F reaches a maximum at $e^{1/e}$ and

that $F'(e^{1/e})$ is infinite. This gives us the amusing but (as far as i know) useless fact that

$$(e^{(1/e)})^{(e^{1/e))(e^{1/e))(e^{1/e))(e^{1/e))(e^{1/e)\cdots}}=e}$$

(Check this: if the LHS is to evaluate to x we must have $(e^{(1/e)})^x = x$ and e is certainly a solution to this equation.)

We can get a power series expansion of F for values of x not much bigger than 1. Let Σ be the power series for $F(1+x)$. Then we have

$$(1+x)^{\Sigma} = \Sigma$$

and we can use the binomial theorem to expand the left hand side. This gives us a sequence of equations expressing later coefficients of Σ in terms of earlier coefficients in a wellfounded way. I haven't worked out the general formula for a_n the coefficient of x^n in $F(1+x)$ tho' in principle it could be done. ($a_0 = 1$ for a start!)

I gather there is an article by Alan Beardon about this. I shall try to track it down.

Analysis and Logic, in their very different ways, come to deal with infinitely long formulæ. These infinitely long formulæ come in two different flavours. From the logician's perspective these two flavours are importantly different; from the Analysts' point of view the one is as bad as the other, both being equally the work of the Devil – *prima facie* at least.

2.2 Taming infinite expressions

A logician would find it natural to make a distinction that to an Analyst probably looks artificial. An Analyst probably thinks that there is no difference between the infinite expression abbreviated as

$$1 + 1/2 + 1/4 + \cdots 1/2^n \cdots$$

and the finite expression $\Sigma_{i\in\mathbb{N}} 2^{-i}$ (tho' the analyst would probably write it $\Sigma_{i=0}^{\infty} 2^{-i}$). For the logician the first is an expression of $L_{\omega_1,\omega}$ and the second is an expression from a first-order language with some second-order parameters – typically $\mathbb{N}$ or some natural subset of it.

I won't dwell on this difference (despite being a logician) beco's it doesn't illuminate the problem of extending the – perfectly satisfactory – semantics for the first-order theory without second-order parameters to the two theories just mentioned. These two theories are in effect equivalent in real life beco's, altho' there are uncountably many infinite sums in the style $1+1/2+1/4+\cdots 1/2^n\cdots$, the only infinite sums that concern us have finite character and can be given finite descriptions.

We have infinite expressions in worse languages than $L_{\omega_1,\omega}$ (continued fractions for one) but the same goes for them. The continued fraction constructor

can be seen logically as a constructor that takes a function $\mathbb{N} \to \mathbb{R}$ (or perhaps $\mathbb{N} \to \mathbb{C}$) and returns a member of $\mathbb{R}$ or $\mathbb{C}$.

HIATUS

Logicians are interested in semantics, the process of assigning meaning. In mathematics (tho' not in ordinary language) the meaning of an expression (once a context has been agreed on) can be computed from the meaning of its subformulæ; indeed it can be computed in no other way. Each task "find the meaning of ϕ" spawns subtasks, one for each subformula – "recursively". We have no objection in principle to infinitely many subtasks being spawned in this way, because [in principle] we might run them in parallel. Infinite time might be needed. However, if the subformula relation is not *wellfounded* containing infinite descending chains or any loops then we have a problem. Recall the predicament of the hapless Liza who is trying to mend the hole in her bucket. She discovers that the endeavour to mend the hole in her bucket spawns a subtask that required her bucket not to have a hole in it in the first place. Even infinite time is of no help to her, since her problem is that the task can never be *started*.

Since the 1960's logic has been fairly relaxed about syntaxes that allow formulæ to be infinitely long. The literature on it is hardly part of the mathematical mainstream, but it is comparatively unproblematic.

$$1 + 1/2 + 1/3 + 1/4 + 1/5 \cdots \tag{2.1}$$

Being a syntax buff, i look at 2.1 and ask myself "Is this an ellipsis?" and i don't mean "is-it-an-ellipsis-because-it's-a-truncation?" I mean "Is it an ellipsis beco's we've left out the brackets?" Addition is associative, so we can leave out brackets. We can leave them out because all results of re-inserting them have the same meaning. Thus any associative operation can be applied to any finite multiset of arguments, so can we apply it to infinite multisets? If we can, then expressions like 2.1 are well behaved. The meaning is obtained by a single infinitary operation.

In contrast if we think of addition strictly as a binary operation, so that 2.1 as an ellipsis for

$$1 + (1/2 + (1/3 + (1/4 + (1/5 \cdots))))\cdots) \tag{2.2}$$

then the problem of assigning meaning to it spawns an infinite descending chain because it requires us to add 1 to

$$1/2 + (1/3 + (1/4 + (1/5 \cdots))))\cdots) \tag{2.3}$$

and evaluating 2.3 is a problem of exactly the same stamp, giving us an infinite descending chain. If we were contemplating

$$1 + (1 + (1 + (1 + (1 \cdots))))\cdots) \tag{2.4}$$

then we wouldn't have an ordinary mere infinite chain, but an actual loop, like poor Liza.

Analysis quite properly regards all infinitary formulæ as problematic, as indeed they are – *prima facie*. The reason why Analysts do not need to distinguish between these two flavours of infinitary formulæ is that their strategy for discovering possible meanings for infinitary formulæ work equally well on both.

Casting my mind back to the 1960s i find that my background assumption was that it was perfectly clear what an infinite sum *meant*; the problem was to *compute* it. Of course what it meant might be `crash` or `fail`. Now – after years of being a logician – it's clear to me that *prima facie* an infinite sum doesn't actually *mean* anything at all; it has to be *given* a meaning. Of course in principle we can do it however we like, but we want to do it in a way that makes for a scar-free junction with the standard semantics for the finitary language. Now quite what does this constraint amount to? We have to respect things like distributivity laws. Of course we do, yes. But why are we right to want this kind of thing?

We say things like "if this infinitary expression means anything at all, it must mean `this`" and we obtain these conditions by expecting the infinitary expressions to obey generalisations of the finitary laws. So – for example – we want the following infinite distributive law:

$$x \cdot \sum_{i \in \mathbb{N}} a_i \;=\; \sum_{i \in \mathbb{N}} x \cdot a_i$$

Notice that, from a logical point of view, there is a clear difference between straightforward infinite sums (terms of $L_{\omega_1,\omega}$ as it were) and expressions like

$$1 + (x + (x^2 + (\ldots))))$$

which have illfounded subformula relation. (for you non-logicians, a formula with illfounded subformula relation is one that kicks off an infinite descending chain of subformulæ in the way that $1 + (x + (x^2 + (\ldots)))$ does). On the face of it this is hugely important beco's an infinite descending chain of subformulæ obstructs the project of giving meaning to an expression by first giving meaning to its subformulæ. This project is called *compositional semantics* by linguists and *recursive semantics* by logicians. When applied to a formula with illfounded subformula relation this process commits one to an infinite sequence of nested calls. A straightforward infinite conjunction is not so problematic. Okay, it spawns infinitely many subtasks but they can all be done in parallel.

However this difference is not as mathematically significant as one might think. One's first thought would be that that all the expressions we encounter that have illfounded subformula relation all have respectable proxies which have an infinitary operation (sum or product) and no infinite descending chain of subformulæ. For example, the associativity of '+' means that

$$1 + (x + (x^2 + (\ldots))))$$

(with its infinite descending chain of subformulæ) has the proxy

$$1 + x + x^2 + \ldots$$

– which has no infinite descending chain of subformulæ – that can be obtained from it by dropping the brackets; and it's OK to drop the brackets beco's + is associative.

However one would be wrong! Think about continued fractions! Nevertheless not much is made of the fact that the subformula relation on continued fractions is illfounded. The value taken by a continued fraction can be represented as the limit of the values taken by a series of finite expressions obtained in a uniform way from the infinite expression

Conditional and Absolute Convergence

Is '$x + y$' the same formula as '$y + x$'? Well, obviously not, but what i mean is: do we think of the dyadic function symbol '+' as taking a *pair* as an argument, or an *ordered pair*? Since + is commutative for *finitary* applications of this syntax it doesn't make any difference, but it matters when we try to generalise to infinitary operators. Is the argument to an infinite sum or infinite product a *stream* of terms or a *set* of terms?

This is where the difference between absolute and conditional convergence comes to life. If the convergence is absolute you can take the argument to be a set; if it's conditional the argument has to be a stream.

> I am quite pleased by the way in which this difference between absolute and conditional convergence – from calculus – turns out to be tied to the difference between two abstract datatypes.

Suppose i want to declare a number system (secretly the reals). It's going to have multiplication $\cdot$ and addition +, both of them commutative and associative, with a multiplicative unit $\mathbb{1}$ and an additive unit 0 and a distributive law $x \cdot (y + z) = x \cdot y + x \cdot z$. We write ' $-x$ ' for the additive inverse of x. Let's also throw in: $x \cdot (-y) = (-x) \cdot y$ as well.

We can choose not to declare what $x \cdot 0$ is; we can declare that it is undefined (unless $x = \mathbb{1}$, when it is defined beco's $\mathbb{1}$ is the multiplicative unit). But when we then change our minds and decide we want it to be defined, there is only one way to go: $x \cdot 0 = x \cdot (y + (-y)) = x \cdot y + x \cdot (-y)$. And if $x \cdot (-y) = -(x \cdot y)$ (which we cunningly assumed with precisely this in mind) then we infer $x \cdot 0 = 0$.

Giving Values to infinite Sums

One decides that the infinite sum evaluates to a limit of the finite sums if that limit exists.

This seems so obvious that one tends to overlook some details.

The first detail is that this allocation is not determined by the semantics of '+'; it requires an actual decision on our part. There is nothing to tell us that an infinite expression has to be given a meaning; we would have been entirely

within our rights to throw up our hands. It's not as if there is a meaning there which we have to nut out, to ascertain.

The second detail is that, in order to justify that allocation, we have to show that the meaning we have decided to give it is the only sensible meaning it could be given. To do this we have to discover what the computational/evaluative behaviour of the sum must be, and then show that the only quantity that behaves properly is the one we have decided on.

We want the semantics to commute with the approximation. There is *syntactic* approximation, and there is *semantic* approximation.

> *We want the semantics for the approximated-infinite-expression to be approximated by the semantics for the finite expressions doing the approximating.*

Of course the fact that it's a sum (and of natural numbers) is a special case. Suppose we have a function f from finite subsets of an index set I taking values in some (complete) ordered set, as it might be $\mathbb{R}$. If f is monotone then we are in with a chance of showing that the infinite sum (or whatever) is defined. We can then show that any two wellorderings σ and τ of a given multiset of inputs converge to the same limit. The point is that every initial segment of σ is included in some initial segment of τ and *vice versa* (because the poset of finite subsets is directed) so that every partial σ-sum is $\leq$ some partial τ-sum and *vice versa*. And the ordering on the ordered set is antisymmetrical.

It is not hard to see that for a multiset to have well-defined sums it must have finite multiplicity.

I should write this out properly: a nice connection between monotonicity and directedness. This covers conditionally convergent series (e.g. alternating harmonic series): one can characterise the difference by saying that the partial-sum function is not monotone!

(Connect this with completion of deterministic monotone blah AC blah)

So we must have a distributivity law for multiplication over infinite sums:

$$x \cdot \sum_{i \in \mathbb{N}} a_i \;=\; \sum_{i \in \mathbb{N}} x \cdot a_i$$

Presumably we want

$$x^{\sum_{i \in \mathbb{N}} a_i} = \prod_{i \in \mathbb{N}} x^{a_i}$$

Presumably these two equations are independent of each other.

There is also the question of making sense of the sum of two infinite sums. What authorises us to interleave?

Can we make sense of $x^{\prod_{i \in \mathbb{N}} a_i}$? I've never seen such a thing, but presumably if the finite exponential "towers" converge to a limit that will be the answer we want.

2.3 My feeble attempts to understand the Riemann zeta function, written up partly to amuse Bruce McKinney and to illustrate the foregoing

Define $\zeta^*(s) =: \zeta(s) \cdot \Gamma(s/2) \cdot \pi^{-s/2}$. (Don't ask why: All Will Be Revealed)
we write '$\overline{x}$' for the complex conjugate of x.
We assert the following without proof, for the moment.

1. $\zeta^*(s)$ is real if s is
2. $\zeta^*(\overline{s}) = \overline{\zeta^*(s)}$
3. $\zeta^*(s) = \zeta^*(1-s)$

Items 2 and 3 give us

$$\zeta^*(1/2+it) = \zeta^*(1/2-it) = \overline{\zeta^*(1/2+it)}$$

whence
$\zeta^*(1/2+it)$ is real.
When real$(s) > 1$, $\zeta(s) = \sum_{n \in \mathbb{N}} \frac{1}{n^s}$
This has the effect that $\zeta(s) = \zeta(s^*)$ (the complex conjugate of s).
There is also a representation as a product.

$$\prod_p \frac{1}{1-(\frac{1}{p})^s}$$

... where the product is taken over all primes. Let $s \to 1$ on the real axis. This shows that there are infinitely many primes.

A bit of calculation that might help

$$\zeta(s) = \sum_{n \in \mathbb{N}} n^{-s}$$

so it's

$$1 + 1/2^s + 1/3^s \ldots$$

which is an expression of infinite length to which – *prima facie* – we cannot give any meaning. This is a general problem. Is there a way of giving a meaning to an infinite sum? Well, one thing we could try is to look at the partial sums – so that, for each n, we add up the first n terms – and see if these partial sums converge to a limit. If they do, we take a deep breath and say that the "sum" of the infinite expression is this thing to which the partial sums converge. We actually have no authority to do it but it doesn't seem to lead to any trouble. There is more one can say about this but i'll leave it at that for the moment.

But what if the finite sums don't converge to a limit, so that we are deprived of an obvious candidate for meaning of the infinite expression? The Riemann ζ function gives us a useful illustration. It's not hard to see that if $s > 1$ then the partial sums converge to a limit (in fact it is sufficient that the modulus, the absolute value $|s|$ of s, be greater than 1, but never mind). However it is quite clear that if $s < 1$ then the terms n^{-s} as n gets larger decrease so slowly that the partial sums do not converge. Bummer. However, consider the following expression

$$(1 - 2^{(1-s)})(1 + 1/2^s + 1/3^s \ldots).$$

On the face of it, this is naughty, beco's one of the multiplicands is an infinite expression that *prima facie* has no meaning. Now we have a multiplicative law $(x + y) \cdot z = (x \cdot z) + (y \cdot z)$ and if x, y and z are legitimate quatities we can use it. This multiplication is guaranteed to work as long as there are only finitely many summands (as in this case – where we have only x and y) and we have no authority to suppose that it works to distribute a multiplication over infinitely many things being summed. So **If** (and it's a big 'if') $1 + 1/2^s + 1/3^s \ldots$ is a legitimate quantity **then** we can invoke the multiplicative law to expand

$$(1 - 2^{(1-s)})(1 + 1/2^s + 1/3^s \ldots)$$

to

$$(1 + 1/2^s + 1/3^s \ldots) \; - \; 2^{(1-s)}(1 + 1/2^s + 1/3^s \ldots)$$

and – again as long as $(1 + 1/2^s + 1/3^s \ldots)$ is OK – we can use multiplication again to expand the second summand to $2(-1/2^s - 1/4^s - 1/6^s \ldots)$, and we rearrange to get

$$1 - 2^{-s} + 3^{-s} - 4^{-s} \ldots$$

where the successive terms are of alternate signs. This means (you might like to check this) that the partial sums now *do* converge as long as $s > 0$, and we get a sensible candidate for a value of the infinite expression. This is good tangible progress, because it means that, despite the fact that the series for $\zeta(s)$ does *not* converge if $0 < s < 1$, the expression for $(1 - 2^{(1-s)})\zeta(s)$ *does* converge to something, and we can recover $\zeta(s)$ from that something by dividing it by $1 - 2^{(1-s)}$. In doing so, we have given a meaning to an infinite expression that *prima facie* didn't have one.

> I now think that the thing to say is this. Take the alternating series (which converges conditionally for $0 < s \leq 1$) sum it and divide by $1 - 2^{(1-s)}$. There must be something dodgy about the execution of the division. Generally one has to pay attention to the intension/extension distinction and to the idea of *evaluation*.

Notice that we haven't proved that $\zeta(s)$ **is** defined when $0 < s < 1$, but we have found something that must be the meaning of $\zeta(s)$ (for such s) if it has a meaning at all. Acting on the assumption that it bears that meaning turns out not to get us into trouble.

> How come making that assumption does not get us into trouble? Presumably this is beco's whenever we try a trick like multiplying the series by

> $1 - 2^{(1-s)}$, or by the number of hairs on Euler's head – or whatever – we do not obtain an answer which conflicts with what we have just obtained. I know no proof of this, and would have no idea whence to obtain one. This is a *lacuna*.

To put it another way, we have found an expression which is defined wherever the formula for $\zeta(s)$ is defined and agrees with $\zeta(s)$ wherever $\zeta(s)$ is defined *and is defined in some other cases as well.*

This is starting to sound as if analytic continuation is to be understood as a case analysis, so that we declare something like:

$$\zeta(s) = \texttt{if } |s| > 1 \quad \texttt{then } \sum_{n \in \mathbb{N}} n^{-s}; \\ \texttt{else } (1 - 2^{(1-s)})^{-1} \cdot \sum_{n \in \mathbb{N}} (-1)^n \cdot n^{-s}$$

and then we have to check that the joins have no wrinkles.

2.3.1 A digression on Analytic Continuation

Just been to a lecture by Mark Gross about Analytic continuation. You say that a pair $\langle U_2, f_2 \rangle$ is a direct analytic continuation of $\langle U_1, f_1 \rangle$ if (forget about the analyticity for the moment) $U_1 \cap U_2$ is nonempty and $f_1 \restriction U_1 = f_2 \restriction U_2$. The domains are open and path-connected for reasons which become clear to us only later. [the path-connectedness works a bit like convexity] When $\langle U_2, f_2 \rangle$ is a direct analytic continuation of $\langle U_1, f_1 \rangle$ we can amalgamate these two by taking the union $U_1 \cup U_2$. Notice that this last set is path-connected as long as U_1 and U_2 are. We want to do as much amalgamation as we can. Gross (and, i think, the tradition) regards this as a binary relation which is symmetric and reflexive but isn't transitive. It seems to me that it's not really a binary relation but a unary higher-order property of sets of pairs $\langle U, f \rangle$. A set of pairs has this property if the domains mentioned in the pairs have nonempty intersection and the functions agree on the intersection (and the same holds for all subsets of that set of pairs).

But what if we have a chain? The problem is that this relation (which we write '$\sim$') is not transitive. Clearly $U_1 \cap U_2 \neq \emptyset$ and $U_2 \cap U_3 \neq \emptyset$ do not imply $U_1 \cap U_3 \neq \emptyset$. But there is another way in which transitivity can fail. Suppose $\langle U_1, f_1 \rangle \sim \langle U_2, f_2 \rangle \sim \langle U_3, f_3 \rangle$. We might have $U_1 \cap U_3 \neq \emptyset$ but have f_1 and f_3 disagreeing somewhere on $U_1 \cap U_3$. This reminds me of the example from PTJ part II logic sheet of the failure of amalgamation of preference relations.

But - hang on! Isn't there a kind of uniqueness theorem for analytic functions that says this can't happen? Suppose $f_1 \subseteq U_1 \times U_1$ and $f_2 \subseteq U_2 \times U_2$ agree on $U_1 \cap U_2$ (thereby defining an analytic function $f \subseteq (U_1 \cup U_2) \times (U_1 \cup U_2)$)

and

$f_2 \subseteq U_2 \times U_2$ and $f_3 \subseteq U_3 \times U_3$ agree on $U_2 \cap U_3$ (thereby defining an analytic function $f' \subseteq (U_2 \cup U_3) \times (U_2 \times U_3)$ but that f and f' disagree on $U_1 \cap U_3$.

Then we find that f and f' are two distinct analytic functions that agree on the domain U_2 and that isn't allowed, so we don't have to worry!

Let's try to describe this situation in 1a DM style.

We consider two binary reflexive symmetric relations on pairs $\langle f, U\rangle$:

(i) $\mathcal{O}$, of *overlapping* (when $U_1 \cap U_2 \neq \emptyset$) and

(ii) $\mathcal{C}$ of *compatibility* when $f_1 \restriction U_1 = f_2 \restriction U_2$. The fact about uniqueness of analytic functions is $(\mathcal{O} \cap \mathcal{C})^2 \subseteq \mathcal{C}$. To be continued

Is there a constructive problem with analytic continuation?

A function defined on the whole of the complex plane by analytic continuation seems to need a case split: "if $x \in A$ do this, if x is in B do this..." and so on. It seems that you need to know whether you are in A or in B. However, this might not be the case. Look more closely...

Suppose f is defined on A but not anywhere else, so that "$(\exists! y)(y = f(x)$" is true iff $x \in A$, and suppose g defined on B similarly. This can be arranged even if *prima facie* f is defined more widely, for we can simply declare $y = f^*(x)$ iff $y = f(x) \wedge x \in A$ ... and hope that f^* is a function too. Suppose further that f and g agree on their intersection. We want to show that

$$(\forall x)(\exists! y)(y = f(x) \vee y = g(x)).$$

So suppose

$$(y = f(x) \vee y = g(x)) \wedge (z = f(x) \vee z = g(x)).$$

We want $z = y$.

Using distributivity we get four cases:

(i) $(y = f(x)) \wedge (z = f(x))$
(ii) $(y = f(x)) \wedge (z = g(x))$
(iii) $(y = g(x)) \wedge (z = f(x))$
(iv) $(y = g(x)) \wedge (z = g(x))$

and all four imply $y = z$. (i) and (iv) work beco's of transitivity of '='; (ii) and (iii) work beco's f and g agree on their intersection.

So the witness y is unique if it exists; but, as long as $x \in A \cup B$, there is one. So analytic continuation is constructive!

2.3.2 We should say a bit about the connection with primes

Consider the expression $(1 - p^{-s})^{-1}$. You know your geometric progressions well enuff to know that this can be expanded as

$$1 + p^{-s} + p^{-2s} + p^{-3s} + \ldots \tag{1}$$

Now consider the product

$$\prod_p (1 - p^{-s})^{-1}$$

over all primes p. Next write out each factor $(1-p^{-s})^{-1}$ as a power series as in (1) above. Now comes the tricky part. Think of of the infinite product as the product of all the series (1) so we are looking at

$$\prod_p 1 + p^{-s} + p^{-2s} + p^{-3s} + \dots$$

the product taken over all primes p.

We are taking the the product of infinitely many things, each of which is an infinite sum. This infinite product is an infinite sum of terms, each of which is itself an infinite product. Each such term (a *summand*) is a product containing one factor from each of the factors in the style (1) above. Let us think about these summands. Each summand is a product of infinitely many things, one from each series in the style (1) above. So it's a product of lots of things like p^{-ks} where s is fixed, p varies over the primes, and k is a natural number that tells you how far along in (1) the factor is.

The intention is to show that, for every natural number n, n^{-s} appears as precisely one of these summands. This will work beco's every natural number is uniquely expressible as a product of prime powers. Let n be – say – $3^{17} \cdot 7^9 \cdot 19$. Then n^{-s} is obtained by picking 3^{-17s} from the third row, 7^{-9s} from the seventh row, 19^{-s} from the 19th row, and 1 from every other row.

That way every number of the form n^{-s} appears as a summand. Indeed each such summand appears as a product in which all but finitely many of the factors are 1. What about the summands which have infinitely many factors that are not 1, but are things of the form p^{-ks}? It's not hard to show that all such infinite products are 0.

This establishes that

$$\zeta(s) = \prod_p (1-p^{-s})^{-1}$$

And that is where the connection between the zeta function and primes comes from.

2.3.3 Afterthoughts

It now occurs to me that the $1-2^{(1-s)}$ was done wrong; we should've done it the other way round. Surely what we want to do is start with the series

$$1 - 2^{-s} + 3^{-s} - 4^{-s} \dots$$

This series is conditionally convergent for $0 < s < 1$. We can then divide by $1-2^{(1-s)}$ to obtain the series for the ζ function. We can do either a long division or we can multiply by the GP/binomial power series for $(1-2^{(1-s)})^{-1}$.

A puzzle for me is: since the sum of the alternating series is good for $0 < s < 1$, and division by $1-2^{(1-s)}$ and multiplication by the infinite series for $(1-2^{(1-s)})^{-1}$ both seem OK why is the result (the +ve series) not

good for $0 < s < 1$? The multiplication/division to which one is committed is infinitary and that can't help. Also the fact that the alternating series is conditionally convergent rather than absolutely convergent must be significant: it means that the manner in which you do the multiplication/division matters. But one can turn the alternating series into an absolutely convergent one by bundling consecutive terms. Should write that out.

The conditionally convergent alternating series can have adjacent terms bundled into an absolutely convergent series which converges for all strictly positive s. We can divide this series by $1 - 2^{(1-s)}$, and we get another series. How can it come about that this other series is not convergent for $s < 1$? The answer is that the process of long division is not to be trusted.

2.3.4 Continued Fractions

$$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}$$

Suppose this infinite continued fraction has a meaning. What might that meaning be? Admittedly there is an infinite descending chain of subformulæ obstructing our search for that meaning, but that very infinite descending chain has a periodic structure which we can exploit. If we suppose that the continued fraction takes the value a, we get immediately the equation

$$a = 1 + \frac{1}{1 + a}$$

from which we can obtain a quadratic equation. We can solve this equation and we find that the root turns out to be the limit of the sequence of initial segments of this infinite continued fraction. Everything fits.

To deal with infinite continued fractions (at least to deal with this one) we don't need infinite sums or infinite products, we just need limits.

What happens if you multiply two continued fractions?

$$\begin{aligned}
&(1 + 1/a) \cdot (1 + 1/b) \\
&= 1 + 1/a + 1/b + 1/(a \cdot b) \\
&= 1 + (a + b + 1)/ab \\
&= 1 + 1/(ab/(a + b + 1)) \\
&= 1 + 1/(1/b + 1/a + 1 \\
&= 1 + 1/(1 + 1/a + 1/b)
\end{aligned}$$

which looks like progress but actually isn't

2.4 Another can of worms

A lot of infinitary expressions can be approached by nonstandard analysis, since the data types of the entities that index the sums/products can look very like natural numbers.

I seem to remember Conway, on his last (and final) visit to Cambridge (nov 2015) saying something along the lines of how a colleague of his thought that approximation beyond all orders (Stirling's expansion) might be illuminated by considering initial segments of these series of nonstandard length.

Here be Dragons.

A talk from Conway on 19/xi/2015

$$n! = \sqrt{2\pi n} \cdot (n/e)^n (1 + \frac{1}{12n} + \ldots)$$

But the series doesn't converge. The first k terms in the series are good for sufficiently large n. If you take $> k$ terms you get a better approximation, but only for larger n than worked with k. For any given n the error you get by using k terms increases with k.

Conway says he got the idea of disjunctive sum from Go.

Chapter 3

Miscellaneous Group Theory

Not sure where to fit this is . . . here? Or in SymmV.tex . . . ?

No transposition can be a square (i think there is a proof in SymmV.tex)

In a finite permutation group every element is either odd (a product of an odd number of transpositions) or even. We can define a transposition as a nonidentity element whose composition with any of its conjugates is of order 1, 2 or 3. Is it true that in any finite group every element is a product of either an even number of "transpositions" or an odd number of "transpositions" but not both?

Notice that in a finite permutation group no element a can be conjugate to ab where b is a transposition. Is it true that in a finite group no element a can be conjugate to ab where b is a "transposition". Does this give us an example of a fmla in the first-order language of group theory which entails infinity?

Is it true that ny two "transpositions" are conjugate?

Simon Wadesley points out that if the order of the group G is not divisible by 2 or 3 then G cannot contain any elements of order 2 or 3 (think: Lagrange) so there are no "transpositions" so *a fortiori* it can't be the case that everything in a finite group is the product of an even or of an odd number of "transpositions"but not both.

Found this on a piece of paper up at the farm. . .

"Every permutation of odd order is a square (think of polygons and diagonals). Let τ be of odd order, so $\tau = \sigma^2$. What can one say about σ?"

First thing is: is σ unique? Where do we get σ from? We process each τ-cycle individually. τ is of odd order, so each cycle is a $(2n+1)$-cycle for some n. We declare the action of σ on a $(2n+1)$-cycle to be τ^k for some k s. t. $2k \equiv 1 \pmod{2n+1}$. Clearly k has to be $n+1$.

So every permutation of odd order is a square in a unique way? No, beco's σ could move things around between τ-cycles. But there certainly is a *canonical* square root. And every square root of τ can be obtained from it by composing

it with a involution that commutes with τ.

Should say a bit more about this. Let τ be a permutation of odd order, and σ its canonical square root as above. σ "is locally a power of" τ. Suppose π is an involution that commutes with τ. Then $\pi\sigma$ is also a square root, as follows.

$$(\pi\sigma)(\pi\sigma) = \pi^2\sigma^2 = \pi^2\tau = \tau.$$

This is because π commutes with any power of τ – and therefore with σ. But for this we do need $\pi^2 = \mathbb{1}$. So the set of square roots of τ isn't simply a coset of the centraliser of τ – which had been my first thought.

Does this generalise? Replace 'of odd order' by 'of order congruent to a (mod b)', and replace 'is a square' by 'is a kth power'. So k would have to divide every number of the form $a + mb$. For example, if every τ-cycle has length congruent to -1 (mod 3) then τ is a cube.

Isla Staden says you can axiomatise noncyclic groups by a scheme that says, for each $n \in \mathbb{N}$, that if $x^n = \mathbb{1}$ then $x^k = \mathbb{1}$ for some $k < n$. Why did i never notice that?? Actually i may have misremembered it. What she probably said (or should have said) is that the complement of the set of finite cyclic groups can be axiomatised by the scheme $(\exists x)(x^n \neq \mathbb{1})$, and clearly not finitely axiomatised.

Balanced Words

A word is balanced iff (tho'rt of as a string) its entries can be permuted to obtain something that is trivially equal to $\mathbb{1}$. Obvious, natural example – a commutator. In fact:

Remark 1 *Every balanced word is a product of commutators*

Proof:

By induction on the length of words. Every "balanced" word is of the form $aXa^{-1}Y$ for some generator a and – balanced – words X and Y. Now $Xa^{-1} = a^{-1}X[X^{-1}a]$, so

$$\begin{aligned} aXa^{-1}Y &= aa^{-1}X[X^{-1}a]Y \\ &= X[X^{-1}a]Y \end{aligned}$$

Now $X[X^{-1}a]Y$ is a product of commutators iff its conjugate $X^{-1}X[X^{-1}a]YX$ is a product of commutators. But this conjugate is $[X^{-1}a]YX$, and $[X^{-1}a]YX$ is a product of commutators iff YX is a product of commutators. Clearly YX is balanced iff $aXa^{-1}Y$ is, and YX is shorter than $aXa^{-1}Y$ so we have the makings of an induction on word length. Every balanced word is of even length. This step shows that a given balanced word w of length $2n + 2$ is a product of commutators as long as a certain balanced word w' of length $2n$ is a product of commutators. But any balanced word of length 4 *is* a commutator.

■

Observe that this construction is effective.

In the basic Fraenkel model the commutator subgroup of the free group on the atoms is not free, or so it is said.

We can also say what it is for two permutations g and h to be powers of the same permutation. You say that either they are identical or they commute and disagree on at least one argument. You don't need naturals!

Can we say in the language of group theory what it is for a permutation to have infinite support?

Suppose we could say that a permutation had only one cycle. (We know how to say that a permutation is a transposition!)

Consider "$(\exists b)(a \cdot a^b)$ is a transposition" That certainly implies that a has infinite support. Trouble is that it implies a lot more besides.

Conjecture 1

Suppose $\sigma \in Symm(X)$ has n-cycles for arbitrarily large finite n.

Then $j(\sigma)$ has infinite cycles.

Proof: .

One wants to pick an n-cycle for each n and then take their union to obtain a subset of X whose $j(\sigma)$-cycle is infinite, but of course that uses AC. It would be good to find a proof that doesn't use AC. For our purposes we would be satisfied with a choice-free proof that, for even one n, $j^n(\sigma)$ had an infinite cycle.

So here is the challenge, posed in language that a group theorist might understand:

The symmetric group on a set X acts in an obvious way on each of the iterated power sets of X: $\mathcal{P}(X)$, $\mathcal{P}^2(X)$ and so on. Suppose X is infinite and σ is a permutation of X that has n-cycles for arbitrarily large n. If we have countable choice then it's obvious that σ has an infinite cycle under the action on $\mathcal{P}(X)$: just pick a representative from each finite cycle: the resulting set of representatives belongs to an infinite cycle. The challenge is to do this without using countable choice. It would be nearly as good to show that such a σ must have an infinite cycle under the action on $\mathcal{P}^n(X)$ for some bigger n – any would do for my purposes. The thought is that $\mathcal{P}^n(X)$ just might have enough structure for countable choice not to be needed.

The more i think about it the more the answer looks like 'no'.

■

Lunch 8/iv/2014

Iddo Rousseau points out that S_n the symmetric group on n elements ($n \in \mathbb{N}$) is a two generator group. The two generators are (i) an n-cycle σ and (ii) a single transposition τ of two elements adjacent in σ. $n-1$ other transpositions are to be had by conjugating τ by some power of σ. Not sure how you get the rest of them.

In consequence every finite group is a [subgroup of a] two-generator group.

3.1 Some nuggets from Zila's talk at ASL 2011 at Welly

Any two finitely generated abelian groups that are elementarily equivalent are iso. Not true for f.g. nilpotent groups: F_2 and F_n are elementarily equivalent for $n \geq 2$ but of course not iso.

Any finitely generated group elem equiv to a free group is hyperbolic;
follows from
Any finitely generated group elem equiv to a hyperbolic group is hyperbolic.

Th(G), G a free group is stable;
Th(G), G torsion-free hyperbolic is stable.

$\forall \phi$ in the language of group theory there is $k \in \mathbb{N}$ s.t. for all groups G and all $m \geq k$ the truth-value $[[G * G * G * \cdots G \models \phi]]$ (free product of m copies of G) is constant!!!

THAT is absolutely totally fucking mind-blowing. It shows that the free product construction really really *really* destroys information.

3.2 Imprimitivity

A Group G of permutations of a set X **acts imprimitively on** X if there is a partition $\mathbb{P}$ of X s.t., for any $g \in G$ and for any piece p of $\mathbb{P}$, g"$p \in \mathbb{P}$. Apparently this terminology goes back to Galois.

Equivalently G acts imprimitively on X if there is an equivalence relation on X which is a congruence relation for every permutation in G. Is this definition in the textbooks? Is it useful?

Any permutation of a fixed set X can be tho'rt of as a binary structure on X – a digraph $\langle \tau, E \rangle$ wherein every vertex has indegree 1 and outdegree 1. The permutations that commute with a permutation τ are precisely the automorphisms of the corresponding digraph $\langle \tau, E \rangle$. The τ-cycles partition X. The centraliser $C(\{\tau\})$ acts imprimitively on X. This centraliser has a normal subgroup consisting of those permutations of X that fix every every τ-cycle setwise. This is going to be $\{\sigma : (\forall x \in X)(\exists n)(\sigma(x) = \tau^n(x))\}$ which is a *bit* like the set of powers of τ but not exactly. It contains every σ that"is locally a power of" τ.

Why is this never commented on by Group theorists?

Wreath Products

Suppose $\mathbb{P}$ is a partition of a set X into pieces all of the same size. Consider the subgroup G of Symm(X) consisting of those g s.t., for every piece $p \in \mathbb{P}$, g"p, too, is piece of $\mathbb{P}$. Clearly G is obtained – somehow – from Symm($\mathbb{P}$) and the symmetric group Symm(p) of any one of the pieces $p \in \mathbb{P}$ (they're all the same group) by some construction. Is this construction the wreath product?

I've now been to a talk by Peter Neumann where all this was explained. And the answer is: Yes, this is indeed the wreath product. G is the wreath product of Symm($\mathbb{P}$) with Symm(p). This is the connection between primitive/imprimitive group actions and wreath products. See, for example, `http://mathworld.wolfram.com/GroupBlock.html`.

Let's look up a definition of wreath product.

DEFINITION 1 *(Cohn: Algebra vol 1 pp 277-8).*
Let X be a set on which A acts. Then elements of the wreath product $B \wr A$ are pairs $\langle a, \lambda \rangle$ where $a \in A$ and $\lambda \in B^X$ (or $X \to B$ if you prefer) with multiplication

$$\langle a_1, \lambda_1 \rangle \cdot \langle a_2, \lambda_2 \rangle = \langle a_1 \cdot a_2, (\lambda_1)^{a_1} \lambda_2 \rangle$$

It seems that $\lambda^a(x) = \lambda(ax)$ and presumably the multiplication of the lambdas in the definiens is pointwise.

G has a normal subgroup consisting of those h s.t. for every piece $p \in \mathbb{P}$, $g``p = p$.

But presumably not every instance of an imprimitive action is a wreath product – the pieces might not all be the same size.

Imprimitivity and Randall's Proof of Con(NF): coming to a near-litter near you

Now suppose X is an infinite set, and $\mathbb{P}$ is a partition of X as before. Consider the subgroup G of Symm(X) consisting of those g s.t., for every piece $p \in \mathbb{P}$, $g``p$ has small difference with some piece of $\mathbb{P}$. Again, $G$ has a normal subgroup, this time the group consisting of those $h$ s.t. for every piece $p \in \mathbb{P}$, there is $p' in \mathbb{P}$ the symmetric difference $(h``p)$ `XOR` p is small.

Not sure how to describe G as a wreath product but never mind.... That may be something worth getting straight.

It seems to me that the κ-ary analogue of this is in play with the permutations Randall wants in his FM models. They act almost-κ-imprimitively on the clans, and the litters are the pieces of the relevant partition.

3.3 Permutations and Øre's theorem

Dear James

Thanks for this. I've started on The first thing you gave me – Øre's thm. He makes free use of weak forms of the axiom of choice without making their use explicit. For example he has the concept of two permutations π and σ (he calls them 'correspondences' but then he is norwegian and the pickled herring has probably addled his brain) being related iff there is a bijection f between the set of π-cycles and the set of σ-cycles such that for all c, $|f(c)| = |c|$. This

is clearly an important notion. He says two such permutations are **conformal**. Do you happen to know if this terminology is standard? (I am going to need a notation for this equivalence relation, because i need to get to the bottom of how much choice one needs to show that *conformal* implies *conjugate*.) If there is such a bijection f then for each π-cycle c we need to pick a member of c and a member of $f(c)$. We pair these two points, and extend this pairing to a bijection between c and $f(c)$. The union of these bijections is a permutation that conjugates σ and π. In doing this we have employed a principle that says that every family of disjoint (finite-or-)countable sets has a choice function. I call this principle GC for Group Choice, co's it's so useful in Group Theory.

(Notice that there is a notion of equivalence weaker than conformality, which is equivalent to conformality if one has enough choice. That is, say σ resembles π iff, for each $i \leq \aleph_0$, σ and π have the same number of i-cycles. This implies that π and σ are conformal as follows:

Consider the family $X = \{X_i : i \leq \aleph_0\}$ where X_i is the set of bijections between the set of π-cycles of size i and the set of σ-cycles of size i. X is countable so we can use countable choice to pick one bijection from each X_i. The union of these bijections is an f witnessing the fact that σ and π were conformal. Observe that the principle we have used here is good old countable choice (AC_ω), not GC. GC gives us choice functions for arbitrary families of countable sets; AC_ω give us choice functions for countable families of arbitrary sets. Not the same)

The lemma on p 308 could be illuminatingly rephrased as:

> (i) Assuming GC, every permutation is conjugate to its inverse.
> (ii) If every permutation is conjugate to its inverse then a permutation is a commutator iff it is the product of two conjugate permutations.

I shall go and look up the 1934 article of Baer that he alludes to; there may be something in there for Nathan and me. It should be in the Betty and Gordon.

The proof of theorem 5 presumably does not use AC. Is the news that every permutation of finite support is a commutator likely to be of any help in the study of HS?

No doubt i shall be troubling you with more of this!
Thanks again!!!
Thomas

The article in question seems to be:
Oystein Ore "Some Remarks on Commutators" Proceedings of the American Mathematical Society **2** No. 2 (Apr., 1951), pp. 307–314. Stable URL: http://www.jstor.org/stable/2032506 .
Thre is a copy in my assorted-paper-archive directory.

Later (mar 2018) every permutation of finite support (in an infinite symmetric group $\mathrm{Symm}(X)$) is conjugate to its inverse – and there is a permutation

of finite support that does the conjugating. But how do we prove that every permutation of finite support is a commutator? Be careful what you wish for. Every commutator is an even permutation in the group of permutations of finite support, so no transposition is a commutator in the group of finite support – tho' it is in the full symmetric group, as follows.

Let s be a permutation consisting of a single cycle. (it doesn't even have to be a transposition). We make countably many pairwise disjoint copies $s_i : i \in \mathbb{N}$ of s^{-1}, all of them disjoint from s. Call this new permutation s^*. It is conjugate to its inverse. The permutation $s \cup (s^*)^{-1}$ (which is of course the same as $s \cdot (s^*)^{-1}$) is conjugate to s^*. This is beco's s is conjugate to s_1 and s_1 is conjugate to s_2, and so on. s is now the product of s^* and $s \cup s^*$ which are conjugates of each other. So s is the product of something with a conjugate of its inverse, so it is a commutator. So every single cycle is a commutator. We can of course do this construction simultaneously for finitely many s, not just one. So any finite product of disjoint cycles is a commutator, which is to say that any permutation with only finitely many cycles is a commutator.

So, in Symm(V), any permutation with only finitely many cycles is a commutator. And every permutation with only finitely many cycles is conjugate to its inverse. The permutations with only finitely many cycles generate a characteristic subgroup.

Actually i think we can do it for any permutation that is conjugate to its inverse; in Symm(V) any permutation that fixes $|V|$-many things and is conjugate to its inverse is a commutator. Aren't these the permutations that Nathan called *flexible*?

Can we find in Symm(V) a permutation that is *not* a commutator?

Can we find in Symm(V) a permutation that is *not* a conjugate to its inverse?

Not if GC is consistent.

Hang on, this is easy. The commutator subgroup is a normal subgroup of small index so it is the whole group, so by Bowler-Forster every element is a product of commutators.

3.4 Some Conversations with Henry Wilton

Today, 11/xi/15, is one of those days when i have too many things to think about.

The family of equivalence classes under finite difference of increasing sequences from $\mathbb{N}$, which i need for the unfolding of frames for TTT. I have to think about injections from one class into another. Is there a canonical one?

Henry (Wilton) has said a number of things to me. The universal cover operation has a universal property, and is idempotent. A universal cover is simply connected. All these spaces are metric. I seem to remember that the Ellentuck topology is not metric; is this beco's it has too many open sets? Must look up exploded graphs again (beco's of edge contractions in Imre's lecture – and is this something again to do with unfoldings? And acyclicity?) And – thinking about the free group on two generators associated with the rose-with-

two-petals – does this presentation arise naturally from a regular grammar? Must think about what the universal cover of the circle (the helix) does to the various reducts, the circular order, the badge relation and so on. (Henry says there are people who have this in hand – but it makes me think again how the topology is the end-result of throwing things away and going higher-and-higher degree). Think about the various covers of the circle that have 2-to-1, 3-to-1, etc maps onto the circle. Where do we get them from? Here is the question i asked Henry

> If Y covers X ($f : Y \twoheadrightarrow X$ cts etc) must it be the case that every $y \in Y$ has a nbhd N s.t $f \restriction N$ is injective..?

Henry sez: yes, absolutely.

Can i think of these covers like the rubber bands that i roll up in things that fit tightly round my finger, like segments of spirals with the ends joined? But aren't these things exactly the same as the circle? So we are talking about 2-to-1, 3-to-1 etc maps from the circle onto itself?

On 27 Mar 2018, at 04:57, Thomas Forster <tf@dpmms.cam.ac.uk> wrote:

Henry, I hope this finds you well. I've been going over my notes, and trying to understand covering spaces. It struck me that, for each natural number n, there is a homomorphism from the circle to itself that is precisely n-to-one. Is there anything one can say about spaces with this property? It presumably says something about the fundamental group...

v best wishes

Thomas

Dear Thomas,

You're right; this is a very natural question, and in fact an active (if slightly niche) topic of research! By the Galois correspondence, the fundamental group G of a space X with this property has the property that, for every n, G has a subgroup G_n of index n isomorphic to itself. It should be difficult for a group G or a space X to have these properties, but no general classification is known yet.

In fact, we had a lectureship candidate this year who works on exactly these kinds of questions. His name is Wouter van Limbeek, and although he hasn't solved the problem completely, he has some very interesting and ingenious partial answers. In this paper – `https://arxiv.org/abs/1710.02179` – he classifies such groups G in the case when the G_n are all normal in G; and in this paper – `https://arxiv.org/abs/1609.06605` – he classifies manifolds X under a similar hypothesis.

The really remarkable thing is that, although as you say these questions are very natural and fundamental to the subject, they haven't really been systematically studied before van Limbeek's work.

I hope that's some help.

All the best,

Henry

Must there be $\tau \in G$ that swaps x and x'? If $G = \mathrm{Symm}(X)$ then the answer is presumably yes. Do we need any choice to prove anything more general?

There may be nothing to worry about. If x and x' belong to the same cycle then a simple rearrangement does it. (It works classically but perhaps not constructively. What about tuples? Something like that will be true but it might be quite hard to state.

3.5 Emily Erlebach on Lagrange

An email from september 2019

Let G be a group and H a subgroup of G.

> LT+: there's a bijection from $H \times (G : H) \longleftrightarrow G$ that maps cosets to cosets
>
> LT: there's a bijection from $H \times (G : H) \longleftrightarrow G$
>
> LT-: the order of H divides G, i.e. there's some set A such that there's a bijection $H \times A \longleftrightarrow G$

LT+ is equivalent to AC: you can argue directly from LT+, but the easiest way to see this: Blass showed that AC is equivalent to 'every group has a transversal'.

LT- doesn't hold in ZF: take a model of ZF with an amorphous group G, and let H be a proper, non-trivial subgroup. H must be finite (otherwise H, gHg^{-1} violate G being amorphous, for some g not in H). If you had a bijection $H \times A \to G$, A would have to be infinite (as G infinite), but then again you would be able to partition G by copies of A.

LT is more complicated. Consider the following choicey axiom, that I'm calling FB (Family of Bijections):

FB: Let I be an set, and $\{X_i | i \in I\}$, $\{Y_i | i \in I\}$ be families of pairwise-disjoint sets indexed by I. Suppose that, for every $i \in I$, there's a bijection $X_i \to Y_i$. Then there is a bijection from $\bigcup_i X_i \to \bigcup_i Y_i$.

FB obviously implies LT, and is obviously implied by AC. But we don't know if FB is equivalent to AC. The most recent paper on this to my knowledge is Higasikawa, 1995 `https://projecteuclid.org/euclid.ndjfl/1040149358`, where an entire hierarchy of choicey statements from AC to the weak partition principle are discussed.

Chapter 4

Pædagogy

Why do people present mathematical induction as "assume it holds for $n = k$; deduce it for $n = k+1$"? Why not: "assume it holds for n; deduce it for $n+1$"?

One of my students is worried about what happens to the totality proof for Ackermann if you permute the arguments. How can i stop him worrying?

4.1 Three Puzzles from Gareth

God bless him.

4.1.1 Doors

You are shown nine doors. Behind one of the doors is a Small Object Of Desire. Each time the bell strikes you are allowed to open one door – any door. If you find the small object of desire the game is over. If you don't, then the small object of desire has to move precisely one step, to an adjacent door. So if it is behind door 3, say, then it must move to door 2 or 4; if it is behind door 1, then it must move to door 2; if it is behind door 9 then it must move to door 8. (No "wraparound"). After the SOOD has had a chance to move the bell strikes again, and you get another chance.

Can you be sure of capturing your Small Object of Desire? If you try thinking about how you do it you can get an estimate of long an algorithm would run.

If there are only three doors (doors 1, 2 and 3) you open door 2 twice.
(Two openings)

What if there are four doors? You open door 2 twice. Then you pause for thought. If you didn't capture the SOOD before the pause, where is it now? It can't be behind door 1, but it could be anywhere else. Now you open door 3 twice. Suppose you don't catch it. *What does this tell you about where it was during the pause?* If it had been behind doors 3 or 4 you would have caught it. So it was behind door 2 during the pause. It has two moves, beco's you opened door 3 twice. The first time it must have moved to door 1 rather than door 3,

o/w you would have caught it. And if it was behind door 1 it can only move to door 2. So you open door 2. Gotcha!

No, try this:

Open door 2. Now the sood is behind 2, 3 or 4. Open 3. Now it is behind 1 or 3. Open 3 again. Now it is behind 2, so you open 2

(Five openings)

What about 5 doors? Rather like four doors, except that after you have opened door two twice and then door three twice you don't know that the SOOD is behind door one; you know only that if it was behind door 2 during the pause then it will be behind door two after four openings. If you don't catch it that way, then you know it was behind doors three, four or five. Well, that's progress.

I think this works: Open door 2. Now it is behind 2, 3, 4 or 5. Open 3. Now it is behind 1, 3, 4 or 5. Open 4. Now it is behind 2 or 4. By symmetry it can't matter which of 2 or 4 you open so try 2. Now it is behind 3 or 5. Open 3. Now it is behind 4.

How about 6 doors?

Open door 2. Now it is behind 2, 3, 4, 5 or 6. Open 3. Now it is behind 1, 3, 4, 5 or 6. Open 4. Now it is behind 2, 4, 5 or 6. Open 2. Now it is behind 3, 4, 5 or 6. This now resembles the case with 4 doors...

later

If you open the second door and don't find the sood, where is it after you have closed the door? Well, it can't be behind door 1, co's the only way it could've got there is from door 2 and it wasn't behind door 2 co's you checked. So it is behind doors 2, 3 or 4. I think it might be an idea to open door 3. If it's there you're happy of course, but if it isn't there where is it now, after you have closed the door again? It *was* behind 2 or 4, but of course it's moved, so it is now behind door 1 or door 3. Nearly there!

Which door do we open now? Door 3 or door 1? To decide which door to open you have to consider what your situation will be after the next 'no' answer. If you open door 1 and it's not there you then infer that it was behind door 3. But where is it *now*? It's behind doors 2 or 4, and you seem to be in the same situation as before. OTOH if you open door 3 and don't get it, then you know that it was instead behind door 1, and then you know that it can now only be behind door 2. So you open door 2 and there it is!

4.1.2 Black and Blue Balls

You are shown 99 boxes, each with black and/or blue balls in it. You are allowed to count the number of black nd blue balls in each box, and are allowed to take 50 of these boxes. Can you be sure of making your choice in such a way that you make off with at least half of the blue balls and *at least half of the black balls?*

For the moment i'm going to assume that the number of blue balls and the number of black balls are both odd. It'll keep things simple and i don't think it'll make much difference.

I spent several months of odd moments beating myself black-and-blue (ha!) with this problem, being too proud to ask for help. After all, this is a problem for first-year students! Then at lunch one day with Gareth Taylor (who it was who blighted my life in the first place by giving me this puzzle) and Matt Saxton, I mentioned this problem, and Matt's immediate reaction was to try to do by *reductio ad absurdum*. (He is an Applied Mathmo and has no Finer Feelings) "Suppose you couldn't ..." he said.

He's right. Let me take up the thread. Suppose, as he says, that you can't. Then, whenever you pick up 50 boxes, you either have a minority of blue or a minority of black. Fair enough, but you can actually say a bit more ... you have a minority of precisely one of the two colours. That's beco's if you didn't, then the other 49 would have a majority and both blue and of black. So the same goes for collections of 49 boxes:

> **If X is a set of 49 boxes, then it contains either more than half the blue balls or more than half the black balls, but not both.**

So every set-of-49-boxes is either Blue (N.B: upper case) or Black. Notice also that

> **If X and Y are two disjoint sets-of-49-boxes then they are of opposite colours**.

Obvious but useful!

But now we notice that if X and Y are two sets-of-49-boxes (or sets-of-49-or-50-boxes) with $|X \texttt{ XOR } Y| < 3$ (i.e., there are at most two boxes that belong to one set but not the other) then X and Y are the same colour. But then you can walk from any set-of-49-boxes to any other set-of-49-boxes while preserving colour. This is clearly absurd. So we have deduced the contradiction we sought.

Now why did I not see that? It's because I will never use a proof by contradiction if I can help it. I was beating myself black-and-blue looking for a direct proof – for example one by induction on the number of boxes. Maybe there is such a proof (there damned well should be!!) but this one will have to do for the moment.

Some further material kindly provided by Gareth

Solution 1

Let X be any set of 50 boxes which wins on blue, and let Y be the other 49 boxes.

If X wins on black, we're done. If not, then Y has at least as much black shared across fewer boxes, so there are boxes $x \in X$ and $y \in Y$ such that y has more black than x.

Consider $Y \cup \{x\}$. This is a 50-set which wins on black (since even if X and Y had equal black before, x has added at least one black to Y), so if it

wins on blue, we're done. Otherwise, $X \setminus \{x\}$ has at least as much blue, so $(X \setminus \{x\}) \cup \{y\}$ wins on blue, and has more black than X did.

Repeat until done. (Or choose X to have maximal black in the first place.)

Solution 2

If boxes X and Y have black$(X) \geq$ black(Y) and blue$(X) \geq$ blue(Y), then we take X and discard Y, and we're done by induction.

Otherwise, for every pair we have black$(X) \geq$ black(Y) and blue$(X) \leq$ blue(Y), or *vice versa*. (One inequality is strict, but this doesn't matter.) So we can order the boxes $X_1, \ldots, X_{99}$ such that the black totals are increasing and the blue totals are decreasing. We then take all the X_{odd}.

This wins on black since black$(X_{99}) \geq$ black(X_{98}), and black$(X_{97}) \geq$ black(X_{96}), ..., and black$(X_3) \geq$ black(X_2), and any tie is broken by black(X_1). And it wins on blue since blue$(X_1) \geq$ blue(X_2), ..., blue$(X_{97}) \geq$ blue(X_{98}), and any tie is broken by blue(X_{99}).

4.1.3 The Moon-base puzzle

Interestingly the slick way to do this was suggested to me by my 1a CS supervisee Maxim Webb (mgcw2). He thought it was something to do with Borsuk's antipodal fixed point theorem, and perhaps in some sense it is, but it can be solved by those ideas without making Borsuk's theorem explicit. The clever part of his idea is: make the problem continuous, and that is what i am going to run with.

You start at some point on the circle, and you inhale petrol as you go clockwise (it doesn't matter, but you have to fix the sense one way or the other) and the inhalation function is of course continuous. You also consume it as you go, but at a constant rate. Let us assume that you inhale precisely enough petrol during a circuit to get back to exactly where you started. Plot against time the amount of petrol you have in your tank. It starts at 0 and is 0 at the end. Keep on going round and round, thereby describing a periodic function. This periodic function is bounded and continuous. (It almost certainly goes negative at some point but we're not going to worry about that for the moment). Quite which periodic function you get will depend on where you start but they'll be time-shifted versions of each other – plus an offset depending on the petrol. What i mean by that is that if i start (with an empty tank) at some point t a radian or so clockwise from t' then i don't get the same function as i would have got starting at t' shifted by a radian or so, it's also lower on the page by the amount of petrol there would in the tank at t had i started with an empty tank at t' instead of at t.

If you start at the wrong place then you might at some point have negative petrol in your tank. Is there a right place to start? Yes. Every point on the curve corresponds to a point on the circle. Start at that point on the circle which corresponds to the least value of this periodic continuous function. Must there be a minimum? Yes: on any closed compact interval this function must

have a minimum. (Every bounded continuous function on a closed compact set has a minimum). Just start at that point!

Now of course the original version is discrete not continuous. Does this matter? No: even in the discrete case, the periodic function defined by the above ruse is continuous. It's piecewise linear and so ...

Some remarks.

(i) We assumed that the moon explorer is travelling clockwise; but the same argument would work if (s)he were travelling anticlockwise. This shows that there is always both a clockwise solution and an anticlockwise solution. Maybe there is a connection here with the antipodal fixed point theorem after all: it's only just occurred to me. For a start it looks to me as if the clockwise solution and the anticlockwise solution are typically going to be distinct. If you are travelling clockwise then the best place to start is at a point where there has been very little petrol anticlockwise of you. And for the anticlockwise solution it's the other way round. Can these two points really be the same?

(ii) This is a *1a* question. The intended victims have not done first-year analysis (yet) and don't know that every bounded continuous function on a closed compact set (OK, a closed interval) has a minimum. Are they supposed to do it without using this fact? Or are they allowed to appeal to an intuitive understanding of it..?

Chapter 5

Coding and Cryptography

5.1 Part II Coding and Cryptography Lectures by Rachel Camina: Notes by Thomas Forster

5.1.1 Lecture I

Coding and Cryptography by Dominic Welsh OUP;
Communication Theory Goldie and Pinch CUP.

$$\text{source} \to \text{encoder} \quad \overset{\text{channel}}{\longleftrightarrow} \quad \text{decoder} \to \text{receiver}$$

A source is a thing that emits messages. Given a source and a channel we wish to be able to transmit messages from the source thru' the channel both economically and reliably and possibly also *privately*. Hence Cryptography!

Economically: common messages have short codes;
Reliably: error detection and correction

(The test for divisibility by 11 detects transposition of adjacent characters.)

A communication channel takes letters from an alphabet and emits letters from another alphabet. The second alphabet is typically the same as the first alphabet (as in the binary symmetric channel below) or at least has significant overlap with it (as in the binary erasure channel below). The thought seems to be that you can marry up the inputs to the channel with the outputs. Typically one hopes that the input emerges the other end unscathed (which is why the two alphabets are nearly the same). Indeed if one could not marry up input packets with output packets one wouldn't have the concept of a channel actually *transmitting* at all.

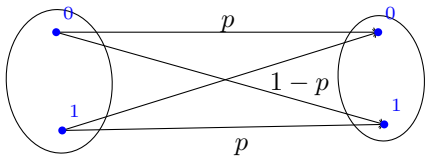
In particular a **Discrete Memoryless Channel** is one for which, for any two characters i and j from Σ, the probability that i is received given that j is

transmitted depends only on i and j and not on the disposition of the planets at the time of transmission.

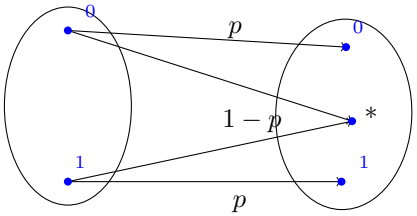
Information about these probabilities is contained in the **channel matrix**; (rows sum to 1)

Examples:

- The **Binary Symmetric Channel** sends 0s to 0s (and 1s to 1s) with probability p, and to the other with probability $1 - p$. The "symmetric" bit is beco's the probability is the same for the two characters.



- The **Binary Erasure Channel** has input alphabet $\{0, 1\}$ and output alphabet $\{0, 1, *\}$. It sends 0 and 1 to $*$ with probability p and transmits them correctly with probability $(1 - p)$. (It never sends 0 to 1 or 1 to 0.) Thus p is the probability that a symbol gets read properly.



Informally the capacity of a channel is the highest rate at which information can be reliably transmitted. There some undefined concepts in there, which is why this definition is informal!

The next section is

I: Noiseless Coding

(with II: Error control codes, and III: Cryptography to follow)

We have an alphabet Σ (echoes of languages and automata!); Σ^* is the set of words over Σ. xy is the concatenation of the two strings x and y; $|x|$ is the length of x. Thus $|xy| = |x| + |y|$.

A **code** is a function $f : \Sigma_1 \to \Sigma_2^*$. Values of f are **codewords**.

Greek Fire Code (208 BC). Σ_1 is the usual Greek alphabet $= \{\alpha, \beta, \gamma, \ldots \omega\}$; $\Sigma_2 = \{1, 2, 3, 4, 5\}$. Used for signalling from hilltop to hilltop. The code sends α to 11, β to 12, $\ldots \omega$ to 54. You transmit a message by holding up a number (≤ 5) of burning torches.

Any function $f : \Sigma_1 \to \Sigma_2$ lifts naturally to a function $f^* : \Sigma_1^* \to \Sigma_2^*$.

We say f is *decipherable* iff f^* is injective. For injectivity of f^* injectivity of f is necessary but not sufficient. Consider $f : \{1, 2, 3, 4\} \to \{0, 1\}^*$ defined by $f(i) =$ binary representation of $(i - 1)$. Evidently f is injective but f^* is not.

An afterthought, a talk with Randall, later

There's a difference between Polish notation and Reverse Polish notation. Polish notation is top-down, so you can tell when you have reached the end of the formula. Reverse Polish notation is bottom-up, and you can't. But the only difference is that one is a reverse of the other!!

5.1.2 Second Lecture

If $f : \Sigma_1 \to \Sigma_2^*$ is injective then it is decodable under any of the following circumstances:

(i) All codewords are the same length. Such an f is a **block** code. Greek fire code is one such;

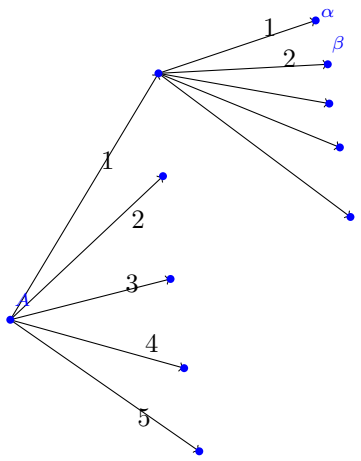
(ii) The code uses one of the characters in Σ_2 as an end-of-word flag. E.g. spaces in our second example above. These are called **comma codes**;

(iii) **Prefix-free** codes: no code word is an initial segment ("prefix") of any other code word.

THEOREM 1 *Kraft's Inequality* Theorem 1.1 in Dr Camina's enumeration

Write 'm' for $|\Sigma_1|$ and 'a' for $|\Sigma_2|$. Then a prefix-free code $f : \Sigma_1 \to \Sigma_2^*$ with word lengths $s_1 \cdots c_m$ can be found iff $\sum_{i=1}^{m} a^{-s_i} \leq 1$.

Proof:



Given the inequality we can

Theorem 1.2 in Dr Camina's numbering

THEOREM 2 *(McMillan) Every decipherable code satisfies Kraft's inequality*

Proof:

Suppose $f : \Sigma_1 \to \Sigma_2^*$ is decipherable, with codewords of lengths $s_1 \cdots s_m$. (Recall that $|\Sigma_1| = m$ and $|\Sigma_2| = n$). Write 's' for $\max\{s_i : i \leq m\}$. For $r \in \mathbb{N}$ consider

$$(\sum_{i=1}^{m} a^{-s_i}) = \sum_{l=1}^{rs} b_l a^{-l}$$

where b_l is the number of ways of choosing codewords with total length l.

Now we invoke decipherability: $b_l \leq |\Sigma_2|^l = a^l$.

Then

$$(\sum_{i=1}^{m} a^{-s_i})^r \leq \sum_{l=1}^{rs} a^i a^{-l} = rs$$

So

$$\sum_{l=1}^{m} a^{-s_i} \leq (rs)^{1/r} \to 1 \text{ as } r \to \infty.$$

■

COROLLARY 1 *A decipherable code with prescribed codeword lengths exists iff there is a prefix-free code with the same codeword lengths.*

Entropy

Wooo!!

A measure of randomness or uncertainty[1].

Consider a random variable X taking values $x_1 \dots x_n$ with probabilities $p_1 \dots p_n$. Then the **entropy** $H(X)$ is (roughly) the expected number of tosses of a fair coin needed to simulate X, or (better) the number of fair (see below) yes/no questions you need to ask to establish any one value of X.

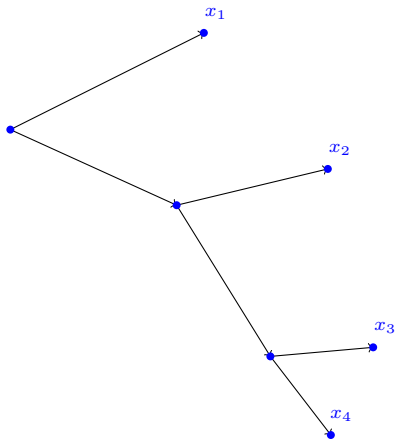
Some examples might help

Example 1:

Four equally likely outcomes, $\{x_1, x_2, x_3, x_4\}$. You ask first whether the answer is in $\{x_1, x_2\}$ or $\{x_3, x_4\}$. Then you know which pair it's in. Then your second question is: which member of that pair? Two questions: entropy is 2. Piece of cake. Well, *two* pieces of cake.

Example 2:

Four outcomes x_1, x_2, x_3 and x_4 with probabilities 1/2, 1/4, 1/8 and 1/8 respectively.



How many questions do we need to ask? Your first question is "Is it x_1?" and only if the answer is "no" do you go on to ask the second question "Is it x_2?", and so on. Half the time it's x_1 and in that case you only ask one question: contribution 1/2. A quarter of the time is x_2 and then you ask two questions: contribution $(1/4) \cdot 2$. One eighth of the time it is x_3 and then you have to ask three questions: contribution $(1/8) \cdot 3$. Finally one eighth of the time it is x_4 and then you have to ask three questions: contribution $(1/8) \cdot 3$. The expected number of questions adds up to 7/4, which is smaller than the expected number of questions in the first example.

[1] Heat is work and work's a curse
And all the heat in the universe
It's gonna *coooool* down as it can't increase.
Then there'll be no more work
And there'll be perfect peace.
Really?
Yeah, that's *entropy*, man!

OK, but how many questions do i need to ask if X takes two values x_0 and x_1 with probabilities 2/3 and 1/3 respectively? Why is that not just *one* question, for heavan's sake? Well yes, it is one question, but it's not a fair binary question. The way through this is pointed to me by my supervisee Jack Rickard. Let's prink out this case with more detail. There is a gremlin that spits out elements of $(0, 1)$; x_1 is the event that the gremlin picks something in $(0, 2/3]$ and x_2 is the event that the gremlin picks something in $(2/3, 1]$. The rules are that you can only ask fifty-fifty questions. So you can't ask "Is it x_1?", you have to ask: "is it less than 1/2?". One half of the time it will be less than 1/2 so you get the answer "yes", infer that it is x_1, and you stop. If you get the answer 'no' you then ask "Is it greater than 3/4?" If you get the answer "yes" you know you are in x_2 and you stop. OTOH if you get the answer "no" you need to press on.

Clearly the chances of your having to ask n questions is 2^{-n} so your expected number of questions is $\sum_{2 \leq n \in \mathbb{N}} n \cdot 2^{-n}$. But that is clearly greater than 1!

The moral is that this identification of the entropy with the number of questions you expect to have to ask doesn't reliably work unless the probabilities are binary (or is the word "dyadic") rationals – denominator a power of 2. It would be nice to understand this a bit better, he says wistfully.

Definition 2 *Entropy*
The **Entropy** $H(X)$ *of the random variable* X *is* $-\sum_{i=1}^{n} p_i \mathtt{log}(p_i)$
– where the logarithm is of course to base 2.

For my own satisfaction i want to check that the two entropies we have calculated above match this definition. I have a suspicious mind. My name is Thomas, after all.

Example 1. There are four terms: each p_1 is 1/4 and $\mathtt{log}(1/4) = -2$ so that 4 times $(1/4)(-2) = -(1/2)$ makes -2 and minus that is 2. So that seems to work.

Example 2. Again there are 4 terms, so we have

$$p_1\mathtt{log}(p_1) + p_2\mathtt{log}(p_2) + p_3\mathtt{log}(p_3) + p_4\mathtt{log}(p_4)$$

Substituting in the actual values we get

$$(1/2)\mathtt{log}(1/2) + (1/4)\mathtt{log}(1/4) + (1/8)\mathtt{log}(p_3) + (1/8)\mathtt{log}(1/8)$$

$$(1/2)(-1) + (1/4)(-2) + (1/8)(-3) + (1/8)(-3).$$

Turned out nice again.

Another example: A biased coin with $p(\text{heads}) = p$. Then $H(p) = H(p, 1 - p) = -p\mathtt{log}(p) - (1 - p)\mathtt{log}((1 - p)$ [which doesn't simplify to anything nice in case you were wondering, as i was] Observe that if $p = 0$ or $p = 1$ then $H(p) = 0$. And that's what you want beco's in those circumstances there is no uncertainty.

Check: if $p = 1/2$ then the entropy is 1.

5.1.3 Third Lecture

I think i have now slain the demons that were assailing me about Huffman's algorithm. So here it is, properly explained.

The output alphabet is {'0','1'} (*i.e.* the characters not the numbers).

We are interested in defining coding functions on alphabets, and we are interested in making them parsimonious. A parsimonious code is one that sends common strings to short codewords.

Now a coding function is something that takes members of Σ as inputs and outputs strings from the output alphabet. However, to say whether or not a coding function is parsimonious we need to consider the probabilities associated with members of Σ. So a function is going to be Huffmann (or not) in relation to a decoration of Σ with probabilities. So we are going to need the idea of *alphabets-decorated-with-probabilities*. An alphabet-decorated-with-probabilities is thus a set of pairs <character, probability> where the probabilities add up to 1.

We are going to define a family of parsimonious coding functions that bear the adjective 'Huffman". And we are going to do it recursively.

Definition 3 *A recursive definition of the class of Huffman codes*

Base case*:*

A coding function defined on a singleton alphabet-decorated-with-probabilities is Huffman iff its value is a single character.

Recursive step

Suppose $f : \Sigma \to \{0,1\}^*$ is Huffman. Suppose Σ' differs from Σ as follows. It has $|\Sigma| + 1$ <character, probability> pairs, of which all but the two least probable are to be found in Σ. Thus $\Sigma \setminus \Sigma'$ has precisely one pair, which we shall call $\langle x, p\rangle$. The two least probable pairs in Σ' are obtained from the pair $\langle x, p\rangle$ by replacing it with two pairs $\langle a, p_1\rangle$ and $\langle b, p_2\rangle$ where a and b are new characters not in Σ, and $p = p_1 + p_2$. Notice that x is not in the alphabet Σ'.

We define $f' : \Sigma' \to \{0,1\}^*$ to be the function that agrees with f on Σ but sends a to $f(x) :: 0$ and b to $f(x) :: 1$.

(Here i am using `ML` *notation for consing/appending)*

Then: if $f : \Sigma \to \{0,1\}^*$ was Huffman, $f' : \Sigma' \to \{0,1\}^*$ is also Huffman.

We claim: Every alphabet-decorated-with-probabilities has a Huffmann code; and: Huffmann codes are optimally parsimonious.

Remark 2 *Every alphabet-decorated-with-probabilities has a Huffman code.*

It may have more than one, of course.

Proof:

What we prove is that

$$(\forall n \in \mathbb{N})(\text{Every } n\text{-sized alphabet-decorated-with-probabilities has a Huffman code}),$$

and we prove it by induction on n.

Suppose true for n, so that every size-n alphabet-decorated-with-probabilities has a Huffman code. Suppose further that Σ' is a size-$(n+1)$ alphabet-decorated-by-probabilities. We seek a Huffman code for Σ'. Let $\langle a, p_1\rangle$ and $\langle b, p_2\rangle$ be the two least probable <character, probability> pairs in Σ'. Let Σ be the alphabet-decorated-by-probabilities obtained from Σ' by replacing $\langle a, p_1\rangle$ and $\langle b, p_2\rangle$ by $\langle x, p_1 + p_2\rangle$, where x is a fresh character not in Σ'. $|\Sigma| = n$ so, by induction hypothesis on n, Σ has a Huffman code f. We now modify f to a Huffmann code for Σ' that sends a to $f(x) :: 0$ and b to $f(x) :: 1$, and otherwise agrees with f.

■

What strikes me about this is that the way the proof actually works is quite different from what one expects, or what is spelt out in the textbooks. It is usually sold to us as a UG on alphabets-decorated-with-probabilities. Let Σ be an arbitrary alphabet-decorated-with-probabilities; then blah blah it has a Huffman code. But it's not that! It's a proof by mathematical induction of a formula that has a quantifier over all alphabets-decorated-with-probabilities. It's not Δ_0!

5.1.4 Example sheet 1

Question 1

Sabine Georgescu makes the point that the reverse of a prefix-free code is decipherable.

Question 2

As SG says, if i concatenate a code with its inverse i get something that isn't decipherable!

Question 4

Equality occurs when $p_2 = p_3$.

Question 5

Question 6

5.2 Some Notes on Professor Körner's Notes

This section is largely messages to myself, dear Reader. You read it – if at all – at your own risk.

TWK says ... you compress the signal as much as you can (remove redundancy). The more you compress it the more it looks like a random sequence.

Then you put back a bit of redundancy in a controlled way to do some error detection and correction. Once you do that your bit stream no longer looks so random – it has detectable structure such as: every 7-bit word has an even number of 1s or something like that.

For two alphabets A and B a **code** is an injection $c : A \hookrightarrow B^*$. Our alphabets are finite of course. A code c is **prefix-free** if $c``A$ is an antichain (in the end-extension ordering)[2]. It is **decodable** if each element of B^* is the c-image of at most one element of A^*. To put it another way: c lifts in a natural way to $c^* : A^* \to B^*$; c is decodable iff c^* is injective. If c is decodable there is a Moore machine $\mathfrak{M}$ over the alphabet B that, on being given a string in B^*, will tell you out loud what was the last coded character from A. And that 'if' is an 'iff'.

Which real, live, codes are decodable? I bet the three-letter codes for airports aren't. How about the one-or-two letter codes for Cambridge colleges? A **K** can only ever mean Kings'; an **H** after an **N** is New Hall, but before a **O** is Homerton.

Exercise 1.6

Consider the problematic second code (problematic beco's Professor K*****r writes it with a superscripted tilde which is **so annoying**):

$$c_2(0) = 0;\ c_2(1) = 01;\ c_2(2) = 011;\ c_2(3) = 111.$$

Observe that '0' only ever appears at the *start* of a code-word. Thus, if you are reading a bitstream, whenever you read a '0' you know that you have just stopped reading a code-word. Also, once you have read three consecutive '1's you know you have reached the end of a word. This directly adresses the problem you have with codes that are not prefix-free, namely not knowing when you have reached the end of a word. Thus c_2^* is injective, which is to say that c is decodable.

It might be helpful to think about how to design a Moore machine[3]

A Mealy Machine is an FSM whose output depends on both the present state and the present input. A Moore machine is an FSM whose output depends on only the present state. that decodes bitstreams according to c_2. I think the significance is that you can design a Mealy Machine but not a Moore machine – at least not straightforwardly. This might be worth making a fuss about.

It may seem like an annoying anomaly that there should be codes c where c^* is injective even tho' c is **not** prefix-free but, but it looks less anomalous if you connect it with the difference between Mealy and Moore Machines.

Here is a Mealy Machine for c_2. It has to have a `start` state, a `fail` state, and states for the various strings that have come in since you last recognised a code word.

[2]Presumably we want $c``A$ to be a maximal antichain under the end-extension ordering ...? Every stream of 0s and 1s has precisely one initial segment in $c``A$?

[3]Ask Wikipædia

If you are in state	and you read	go to state	and say
`start`	0	0	
`start`	1	1	
`fail`	0	`fail`	"`fail`"
`fail`	1	`fail`	"`fail`"
0	0	`start`	"0"
0	1	01	
1	0	`fail`	"`fail`"
1	1	11	
11	0	`fail`	"`fail`"
11	1	`start`	"3!"
01	0	0	"1!"
01	1	`start`	"2!"

A discussion answer to Prof. K's Exercise 2.5

How are we to assign (bitstring) codes to the messages
m_1 of probability 1/45;
m_2 of probability 2/45;
m_3 of probability 3/45;
m_4 of probability 4/45;
m_5 of probability 5/45;
m_6 of probability 6/45;
m_7 of probability 7/45;
m_8 of probability 8/45;
m_9 of probability 9/45;
??

Huffman's algorithm requires us to *aggregate* the two least likely messages, and start again. So the set of messages that we are going to encode is not $\{m_i : 1 \leq i \leq 9\}$ but is

$$\{m_1 \oplus m_2\} \cup \{m_i : 3 \leq i \leq 9\}.$$

$m_1 \oplus m_2$ is a message that crops up with probability $1/45 + 2/45$. If you want to **not** think about what $m_1 \oplus m_2$ is then you can think of it as some otherwise anonymous message which features in a problem to which we are reducing the given problem, and about which you know only that it is of probability 1/15. **Anyway** we now rearrange this new set (of eight messages) in order of probability and, again, aggregate the two most improbable messages, namely $m_1 \oplus m_2$ and m_3, so we now have a set of **seven** messages, namely

$$\{(m_1 \oplus m_2) \oplus m_3\} \cup \{m_i : 4 \leq i \leq 9\}.$$

Again we aggregate the two least probable messages, which – since $(m_1 \oplus m_2) \oplus m_3$ has probability 6/45 – are m_4 and m_5. So the coding challenge is now for the set

$$\{(m_1 \oplus m_2) \oplus m_3,\, m_4 \oplus m_5\} \cup \{m_i : 6 \leq i \leq 9\}.$$

The two least probable messages now are m_6 and m_7, so we aggregate them to have the challenge of coding

$$\{(m_1 \oplus m_2) \oplus m_3,\, m_4 \oplus m_5,\, m_6 \oplus m_7,\, m_8,\, m_9\}.$$

The probabilities associated with these five messages are 6/45, 9/45, 13/45, 8/45 and 9/45 respectively, so the two least probable messages are the first and the penultimate, so we aggregate *them* getting

$$\{(m_1 \oplus m_2) \oplus m_3) \oplus m_8,\, m_4 \oplus m_5,\, m_6 \oplus m_7,\, m_9\}.$$

which have probabilities 14/45, 9/45, 13/45 and 9/45 respectively; so we aggregate the second and the fourth to get

$$\{(m_1 \oplus m_2) \oplus m_3) \oplus m_8,\, (m_4 \oplus m_5) \oplus m_9,\, m_6 \oplus m_7\}.$$

which have probabilities 14/45, 18/45, and 13/45 respectively. Finally we aggregate the first and the third to obtain

$$\{(((m_1 \oplus m_2) \oplus m_3) \oplus m_8) \oplus (m_6 \oplus m_7),\, (m_4 \oplus m_5) \oplus m_9\}.$$

Now we finally have two messages, namely

$$(((m_1 \oplus m_2) \oplus m_3) \oplus m_8) \oplus (m_6 \oplus m_7) \text{ and } (m_4 \oplus m_5) \oplus m_9,$$

which we encode as 0 and 1 respectively.

Consequently we encode
$((m_1 \oplus m_2) \oplus m_3) \oplus m_8$ as 00 and
$m_6 \oplus m_7$ as 01;
$(m_1 \oplus m_2) \oplus m_3$ as 000 and
m_8 as 001;
$m_1 \oplus m_2$ as 0000 and m_3 as 0001.

Coding $m_6 \oplus m_7$ as 01 tells us to code
m_6 as 010 and m_7 as 011.

Coding $(m_4 \oplus m_5) \oplus m_9$ as 1 tells us to code
$m_4 \oplus m_5$ as 10 and m_9 as 11,
whence we code
m_4 as 100 and m_5 as 101.

The following table indicates how less probable messages get longer codes.

Message	Probability	Code
m_1	1/45	00000
m_2	2/45	00001
m_3	3/45	0001
m_4	4/45	100
m_5	5/45	101
m_6	6/45	010

m_7	7/45	011
m_8	8/45	001
m_9	9/45	11

Those $\oplus$-words that crop up in the execution of Huffman's algorithm crop up as names of states in the Moore machine that decodes a stream of code words.

A discussion answer to Prof. K's Exercise 3.1

A prefix-free code is an antichain in the perfect binary tree. Decorate each node in the perfect binary tree with the rational number 2^{-n} where n is its distance from the root. The weight of an antichain is the sum of the decorations. Every antichain can be extended to a maximal antichain, and the weight of any maximal antichain in precisely 1. Think about modifying an antichain by replacing a node by its two children, or replacing two sibling inhabitants by their parent.

Prof Körner's Exercise 3.4:

C_4 is coded by 1
C_3 is coded by 00
C_1 is coded by 011
C_2 is coded by 010

Then consider the Moore machine

If you are in state	and you read	go to state	and say
`start`	1	C_4	"C_4!"
C_1	1	C_4	"C_4!"
C_2	1	C_4	"C_4!"
C_3	1	C_4	"C_4!"
C_4	1	C_4	"C_4!"
`trans1`	1	`trans2`	
`trans2`	1	C_1	"C_1!"
`start`	0	`trans1`	
C_1	0	`trans1`	
C_2	0	`trans1`	
C_3	0	`trans1`	
C_4	0	`trans1`	
`trans1`	0	C_3	"C_3!"
`trans2`	0	C_2	"C_2!"

This is a Moore machine so the fourth column depends solely on the third column.

Exercise 4.4

$x \mapsto e^x$ is monotone increasing so

$\texttt{log}(t) \leq (t-1)$ iff

$t \leq e^{(t-1)} = t + t^2/2! + t^3/3! \ldots$

Definition 4.7

Presumably there is no objection to $\mathcal{A}$ being empty...? Any random variable would have Shannon entropy 0. And – from the point of view of this definition at any rate – there is no reason to assume $\mathcal{A}$ to be finite. All that is required is that it should be possible to add up $|\mathcal{A}|$-many reals to get something sensible if those reals are all suff small. So $\mathcal{A}$ can be countable. Come to think of it, how many reals *can* one sum? Well, you certainly can't *ever* sum $\aleph_1$ reals, co's you order them in order-type ω_1 and consider the (increasing) sequence of partial sums. But can you prove – *without* using countable choice – that you can *never* sum continuum-many reals...?

If AC_ω fails and there are infinite Dedekind-finite sets of reals (any such will have a countable partition) is it possible that such a set will have a well-defined sum?

Anyway, let's try to understand "Random variable taking values in $\mathcal{A}$". I think we characterise a random variable by the probability with which it emits each value. Does this identfy a random variable uniquely? If i have two random variables that pick members of $\mathcal{A}$ with the same probability are they the same random variable?

f is a random variable taking values in $\mathcal{A}$ iff

(i) $f : \mathcal{T} \to \mathcal{A}$ where $\mathcal{T}$ is the set of **trials**.

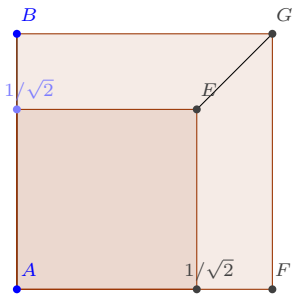
(ii) An **event** is an ordered pair $\langle i, f(i)\rangle$. Thus if I is a (finite) subset of $\mathcal{I}$ then $f \restriction I$ – aka f restricted to I – is a set of events.

(iii) there is a function $p_f : \mathcal{A} \to \mathbb{R}$ telling us how much f likes each member of $\mathcal{A}$, in the sense that if $I \subseteq \mathcal{I}$ is finite and $a \in \mathcal{A}$ then you expect a proportion $p_f(a)$ of the events in $f \restriction I$ to be events in which a happens, i.e., the second component of the event is a. Need a bit of measure theory here.

OK i think i know what Shannon entropy is. $\mathcal{A}$ doesn't even have to be finite. So each $\mathcal{A}$-valued random variable has a Shannon entropy. There are presumably operations on $\mathcal{A}$-valued random variables. How do these operations correspond to operations on the entropies associated with those random variables?

Exercise 4.13

Trubshaw should use Huffman's algorithm to build a tree, and then walk out along that tree by asking questions.



How many questions do you expect to have to ask if you want to know whether your random variable that emits things in $(0, 1)$ has emitted something below $1/\sqrt{2}$? Well it'll take half as many as it takes to answer *two* of these questions simultaneously. Let n be the expected number of questions that we have to ask in order to nail down *one* value of the random variable.

So you ask whether or not both events are less than $1/\sqrt{2}$. Half the time you get the answer 'yes' so that contributes $1/2$ to the expected number of questions. The other half you have to go on to ask which of the two trapezoidal regions you land in. The answer to this will nail down one of the values so all that has to be done is ascertain the other. That will take n further questions, so $n + 1$ questions.

This gives us $2n = 1/2 + 1/2 \cdot (1 + n)$, whence $n = 2/3$.

That does seem a bit low, but that's not really the problem. The problem is that it can't be the same as the entropy, since the entropy has a $\texttt{log}(\sqrt{2} - 1)$ term, and that means it can't be rational.

5.3 A discussion answer to Prof. Körner's Exercise 2.5

How are we to assign (bitstring) codes to the messages
m_1 of probability $1/45$;
m_2 of probability $2/45$;
m_3 of probability $3/45$;

m_4 of probability 4/45;
m_5 of probability 5/45;
m_6 of probability 6/45;
m_7 of probability 7/45;
m_8 of probability 8/45;
m_9 of probability 9/45;
??

Huffman's algorithm requires us to *aggregate* the two least likely messages, and start again. So the set of messages that we are going to encode is not $\{m_i : 1 \leq i \leq 9\}$ but is

$$\{m_1 \oplus m_2\} \cup \{m_i : 3 \leq i \leq 9\}.$$

$m_1 \oplus m_2$ is a message that crops up with probability $1/45 + 2/45$. And what *is* this message, one might ask? Well, it's either m_1 or m_2 but you don't know which when you transmit it. The Linear Logic people will probably have something to say about this. **Anyway** we now rearrange this new set (of eight messages) in order of probability and, again, aggregate the two most improbable messages, namely $m_1 \oplus m_2$ and m_3, so we now have a set of **seven** messages, namely

$$\{(m_1 \oplus m_2) \oplus m_3\} \cup \{m_i : 4 \leq i \leq 9\}.$$

Again we aggregate the two least probable messages, which – since $(m_1 \oplus m_2) \oplus m_3$ has probability 6/45 – are m_4 and m_5. So the coding challenge is now for the set

$$\{(m_1 \oplus m_2)) \oplus m_3,\, m_4 \oplus m_5\} \cup \{m_i : 6 \leq i \leq 9\}.$$

The two least probable messages now are m_6 and m_7, so we aggregate them to have the challenge of coding

$$\{(m_1 \oplus m_2) \oplus m_3,\, m_4 \oplus m_5,\, m_6 \oplus m_7,\, m_8,\, m_9\}.$$

The probabilities associated with these five messages are 6/45, 9/45, 13/45, 8/45 and 9/45 respectively, so the two least probable messages are the first and the penultimate, so we aggregate *them* getting

$$\{((m_1 \oplus m_2)) \oplus m_3) \oplus m_8,\, m_4 \oplus m_5,\, m_6 \oplus m_7,\, m_9\}.$$

which have probabilities 14/45, 9/45, 13/45 and 9/45 respectively; so we aggregate the second and the fourth to get

$$\{((m_1 \oplus m_2)) \oplus m_3) \oplus m_8,\, (m_4 \oplus m_5) \oplus m_9,\, m_6 \oplus m_7\}.$$

which have probabilities 14/45, 18/45, and 13/45 respectively. Finally we aggregate the first and the third to obtain

$$\{((m_1 \oplus m_2) \oplus m_3) \oplus m_8) \oplus (m_6 \oplus m_7),\, (m_4 \oplus m_5) \oplus m_9\}.$$

Now we finally have two messages, namely

$$(((m_1 \oplus m_2)) \oplus m_3) \oplus m_8) \oplus (m_6 \oplus m_7) \text{ and } (m_4 \oplus m_5) \oplus m_9,$$

which we encode as 0 and 1 respectively.

Consequently we encode
$((m_1 \oplus m_2)) \oplus m_3) \oplus m_8$ as 00 and
$m_6 \oplus m_7$ as 01;
$(m_1 \oplus m_2) \oplus m_3$ as 000 and
m_8 as 001;
$m_1 \oplus m_2$ as 0000 and m_3 as 0001.

Coding $m_6 \oplus m_7$ as 01 tells us to code
m_6 as 010 and
m_7 as 011.

Coding $(m_4 \oplus m_5) \oplus m_9$ as 1 tells us to code
$m_4 \oplus m_5$ as 10 and
m_9 as 11, whence we code
m_4 as 100 and
m_5 as 101.

The following table indicates how less probable messages get longer codes.

Message	Probability	Code
m_1	1/45	00000
m_2	2/45	00001
m_3	3/45	0001
m_4	4/45	100
m_5	5/45	101
m_6	6/45	010
m_7	7/45	011
m_8	8/45	001
m_9	9/45	11

5.4 Frames

Look up comma-free codes and circular (cyclic) codes. Arques/Michel codes 1998:

`https://www.sciencedirect.com/science/article/pii/S0898122107005561`

A *frame* is a choice of a starting point. Comma-free codes are nice in that you can always find the correct frame. But you may be able to find which frame you are in even if the code is not a comma code. If you have a long enough string of encoded words you can get ensemble information about the stuff you recover by decoding. The genetic nucleotide code is not comma-free, but you can tell which frame you are in. It's a code of length 3, so there are precisely three frames.

Crick's hypothesis was that the genetic base code is comma-free but apparently it's not true.

5.5 A Christmas Cracker from the Farm, December 2019

Q: "What do ghosts eat?"
A: "Spooketti"

No, but seriously. This came out of one of the crackers at Christmas lunch.

I quote:

"The complete set consists of 6 cards, printed with a series of numbers. Show all the cards to a friend and ask him or her to select one number from any one card. Show all the other five cards to your friend asking him or her to say whether or not the number appears on these cards. Take all the cards on which your friend says the number appears, add together the top left-hand corner number of each card; the total is the number your friend selected."

1	3	5	7	9	11	13	15
17	19	21	23	25	27	29	31
33	35	37	39	41	43	45	47
49	51	53	55	57	59	61	63

2	3	6	7	10	11	14	15
18	19	22	23	26	27	30	31
34	35	38	39	42	43	46	47
50	51	54	55	58	59	62	63

4	5	6	7	12	13	14	15
20	21	22	23	28	29	30	31
36	37	38	39	44	45	46	47
52	53	54	55	60	61	62	63

8	9	10	11	12	13	14	15
24	25	26	27	28	29	30	31
40	41	42	43	44	45	46	47
56	57	58	59	60	61	62	63

16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

Explain what is going on.

Just what in God's name is going on?

The numbers in the top left-hand corner are all powers of two which looks highly suggestive.

Actually it is quite easy. Index the set of cards by $\{1, 2, 4, 8, 16, 32\}$. Every natural number < 64 is expressible as a sum of a unique subset of the index set. So you write the numeral 'n' on every card pointed to by a member of the unique subset of the index set which expresses n.

If you friend chooses 0 it still works! It might be worth trying to find something to say about why there are the same number of numerals on each card.

Chapter 6

Miscellaneous Topology

I seem to remember that the Ellentuck topology (which contains more open sets that the usual topology on $\mathbb{N}^\omega$) is not metric. Zachiri even wrote out a proof for me, God bless 'im. Now what can one say about the family of metric topologies on a fixed set? Is it intersection-closed?

Let X be a set, and let $\mathcal{T}_1$ (with metric d_1) and $\mathcal{T}_2$ (with metric d_2) be metric topologies on X. What about $\mathcal{T}_1 \cup \mathcal{T}_2$ (which is literally $\{t_1 \cup t_2 : t_1 \in \mathcal{T}_1 \wedge t_2 \in \mathcal{T}_2\}$. How do we obtain a new metric from d_1 and d_2?

Every separable metrisable topology Hausdorff, regular, second countable
finer than hausdorf implies hausdorff. Urysohn metrisation theoem.
Look at Bourbaki Topologie Générale
Fix a topology $\mathcal{T}$.

Whenever O is an open set, take $\{C : C \text{ is closed and } C \cap O \neq \emptyset\}$ to be a basic open set in a topology on the set of closed sets of $\mathcal{T}$. This is the *Vietoris Topology*. It is T_1 iff $\mathcal{T}$ is. (Exercise)

For our purposes i think this matters beco's of the Malitz construction of models of Positive set theory. Must check. I asked Isaac about it and he replied that 'Vietoris topology' rings no bells.

What is a closed set? One that contains all its accumulation points. What is an accumulation point of X? It is an x s.t. every open nbhd of x meets X.

The complement of an open set is closed and the complement of a closed set is open? Suppose X is an open set. Consider the interior of the complement of X. Call it I. What are the accumulation points of I? We want them to be precisely the things not in X. If x is an accumulation point of I that *is* in X then it has an open set round it that $\subseteq X$. But I is the union of all open sets disjoint from X so that cannot happen. So all accumulation points of X lie in I. The other direction is easy. So the complement of any open set is closed.

What about the complement of a closed set? Let X be closed, so it contains all its accumulation points. Then nothing in its complement is an accumulation point. So, if $y \notin X$ there is an open nbhd of y that is disjoint from X. But then

the union of all open nbhds obtained in this way is an open set that exhausts everything not in X.

6.1 Circles and Helices

On 27 Mar 2018, at 04:57, Thomas Forster wrote:

Henry, I hope this finds you well. I've been going over my notes, and trying to understand covering spaces. It struck me that, for each natural number n, there is a homomorphism from the circle to itself that is precisely n-to-one. Is there anything one can say about spaces with this property? It presumably says something about the fundamental group.....

v best wishes

Thomas

Dear Thomas,

You're right — this is a very natural question, and in fact an active (if slightly niche) topic of research! By the Galois correspondence, the fundamental group G of a space X with this property has the property that, for every n, G has a subgroup G_n of index n isomorphic to itself. It should be difficult for a group G or a space X to have these properties, but no general classification is known yet.

In fact, we had a lectureship candidate this year who works on exactly these kinds of questions. His name is Wouter van Limbeek, and although he hasn't solved the problem completely, he has some very interesting and ingenious partial answers. In this paper — `https://arxiv.org/abs/1710.02179` `https://arxiv.org/abs/1710.02179` — he classifies such groups G in the case when the G_n are all normal in G; and in this paper — `https://arxiv.org/abs/1609.06605` he classifies manifolds X under a similar hypothesis.

The really remarkable thing is that, although as you say these questions are very natural and fundamental to the subject, they haven't really been systematically studied before van Limbeek's work.

I hope that's some help.

All the best,

Henry

A cts surjection irrationals $\twoheadrightarrow$ reals

There is a cts surjection irrationals $\twoheadrightarrow \mathbb{R}$. Write out the irrational in base 2 and take every second bit. This is onto!!

Thanks to Kris Cao. But then, as Randall says, a top space is compact metric iff it is a hom image of Cantor space. What about hom images of the irrationals?

David Preiss on Convex Borel sets

This is a famous conjecture of Larman and Rogers, proved (using analytic sets) by David Preiss:

We know you get Borel sets by taking open sets and allowing ctble unions and complements.

To get convex Borel sets, is it enough to start with convex open sets and allow the operations of ctble union and ctble intersection only (ie. no complements).

Imre

PS. Of course, transfinitely often, so one means: is the smallest class of sets closed under ctble union and intersection and containing the convex opens also s.t. it contains the convex Borels?

PPS. OOPS: I mean only nested ctble union, of course (to stay in world of convex). So we can say only nested things allowed.

6.2 The Universal Seperable Metric Space: Notes based on a conversation with Randall Holmes

Gentlemen,

The following is the result of a conversation with Randall Holmes, who in an earlier life was a topologist – he did his Ph.D. on this stuff. I am writing it up primarily for my own benefit, and secondarily beco's i suspect it might be pædogogically useful. I am sending this embryo to you beco's (i) if you don't already know it it might amuse you; and (ii) beco's if you do already know it you might have some helpful comments. I am attracted to it beco's – it seems to me that – the typical connections from Model theory are to Algebra and Combinatorics, whereas this Fraïssé construction results in *topological* fun.

Consider finite digraphs, with their edges decorated by **rationals** in a manner that respects the triangle inequality. Observe that two decorated graphs A and B not only have an amalgamation but – if they have nonempty intersection – a *canonical* amalgamation. How are we to decorate edges in $A \cup B$? The only edges we don't know how to decorate are edges (a, b) with $a \in A \backslash B$ and $b \in B \backslash A$. The triangle inequality requires that $d(a, b) \leq \min\{d(a, c) + d(c, b) : c \in A \cap B\}$ and this quantity is of course a rational number and can be used as the value for $d(a, b)$. There is a lower bound as well, beco's $d(a, b)$ cannot be less than $|d(a, c) - d(c, b)|$ for any $c \in A \cap B$.

These decorated digraphs form a family that meets all the conditions required by the Fraïssé construction, so there is a direct limit, which is a strongly homogeneous countable structure with a metric. We complete this structure to obtain a seperable metric space. For the usual Fraïssé reasons this structure is strongly homogeneous, and of course every seperable metric space embeds into it.

6.3 A Lecture by TWK on Met-and-Top Easter Term 2014

. . . reminds me that $\mathbb{R}$ (the real lime) is green and homeomorphic to $(0, 1)$ but one is complete and the other one isn't.

I have just been to Tom K's course on metric and topological spaces. As Tom says, this is ground that has, over the years, been well gone over by lots of very clever people, with the result that the treatment now available is highly optimised. The point is not that highly optimised systems fail catastrophically (true tho' that is); the point is rather that you can have a perfectly satisfactory understanding of your path through the Great Grimpen Mire but still be as lost as the next man once you stray into the mire itself. Having been to a course like Tom's and mastered all the details so that you are on Stapleton's path and know how to stay on it doesn't in the least equip you to help people who have strayed off it and got lost in the mire. "I wouldn't start from there, i'd start from over here" is not the kind of advice they are looking for.

One point worth making (it seems to me) is that *huge* mileage can be made of the fact that the property of being a topology is intersection-closed. If a topology on a set is a special kind of subset of its power set then an arbitrary intersection of topologies is another topology. This explains why we can talk of a topology being generated by a collection of open sets, a *basis*: the topology generated by a given basis is simply the intersection of all topologies that are supersets of that basis. The subspace topology on $Y \subseteq X$ is the least topology on Y making the inclusion map continuous. Any topology on Y that makes the inclusion embedding continuous must contain $(\mathrm{id}_Y)^{-1}$"X' for any open $X' \subseteq X$. And $(\mathrm{id})_Y^{-1}$"X' is just $Y \cap X'$. So the subset topology must be the topology generated by all the $Y \cap X'$ with X' open. I'd always known this was the definition, but i'd never known why!

Fine and Coarse

The more open sets a topology has, the *finer* it is.

If $\{Y_i : i \in I\}$ is a family of topologies, X is a set, and there are maps $f_i : i \in I$ from X to Y_i, then the product topology on $\prod_{i \in I} Y_i$ is defined to be the least ("coarsest") topology making all the projection maps continuous. 'Coarser' means 'fewer open sets' so the coars**est** topology such that blah should be the intersection of all topologies such that blah (always assuming that the intersection is still blah of course!) Let's see what this actually amounts to. Suppose U is an open set in Y_i, then

$$\{f \in \prod_{i \in I} Y_i : f(i) \in U\} \qquad (**)$$

must be an open set. So the product topology on $\prod_{i\in I} Y_i$ is the intersection of all topologies on $\prod_{i\in I} Y_i$ that contain all the sets $**$. But that is simply to say that it is the topology generated by the sets $**$.

These definitions rely on the fact that "projections are continuous wrt $\mathcal{T}$", and so on are intersection-closed properties of topologies.

It will be worth writing this out in detail.

Tom sez the definition of quotient topology is nasty. The [more general] idea is this. Suppose i have a set X equipped with a topology, and a set Y, not yet so equipped. I also have a set F of functions $X \to Y$ and i'm going to call them all continuous, just beco's i happen to feel like it. There is now a *finest* topology on Y that makes all the functions in F cts. 'Finest' is nice, beco's the more open sets there are $\subseteq Y$ the harder it is for any one function $X \to Y$ to be cts.

Mind you, wikipædia sounds quite different!

As Gareth says, with the definition of quotient topology and subset topology one is the finest and the other is the coarsest, and that's beco's they are on different sides of the arrow!

Oh yes *and another thing*.... In the definition of homotopy, it should be slightly easier to state things clearly if one curries the homotopy function $H : X \times [0,1] \to Y$ into $H : [0,1] \to (X \to Y)$. The insistence that the uncurried H be continuous of course goes over to an insistence that the curried H be continuous, and this of course puts a constraint on the topology of $X \to Y$.

Gareth says: you want two spaces X and Y that are homotopic but not homeomorphic. Well, just take those two letters!

6.4 Tikhonov's theorem for Hausdorff Spaces is equivalent to the Prime Ideal Theorem

tf to ptj

Peter,

Does STONE SPACES contain a proof of the equivalence of PIT and Tych for Hausdorff spaces? The notes to one of the chapter points the reader to Los-Ryll-N FM 1951 (a paper which remarkably appears to be in English) but your otherwise estimable and terrifying book has no index. What is the easiest way to learn a proof of this equivalence?

Thomas

ptj to tf

There is a proof in Stone Spaces, but it's mostly in the notes rather than the main text: see Remark III 1.10 on page 90 and the notes on pages 119 and 120.

However, the way I deduce (PIT $\rightarrow$ Tychonoff), via the fact that Tychonoff for locales can be proved without choice, is not the standard one (and certainly not the original one of Los–Ryll-Nardzewski): I forget exactly how they did it, but the easiest "classical" way is to use PIT to prove that "compact Hausdorff" is equivalent to "every ultrafilter has a *unique* limit", and then to observe that the latter property is (without choice) inherited by products. (See also Wistar Comfort's 1968 paper cited in Stone Spaces.)

Peter

$\mathcal{O}(\mathcal{T})$ is the collection of open sets of $\mathcal{T}$.

Say a point x is a **convergence point for** an ultrafilter $\mathcal{U}$ iff every open neighborhood of x is in $\mathcal{U}$.

Remark 3 *A space is Hausdorff iff every ultrafilter has at most one convergence point.*

Proof:

$L \rightarrow R$

Suppose $\mathcal{U}$ has two convergence points a and b. Then there are disjoint open neighbourhoods U_a round a and U_b round b, and they can't both belong to $\mathcal{U}$, being disjoint. So if $\mathcal{T}$ is Hausdorff no ultrafilter on $\mathcal{T}$ can have more than one convergence point.

$R \rightarrow L$

Suppose no ultrafilter on $\mathcal{T}$ has more than one convergence point, and suppose that a and b are counterexamples to the assertion that $\mathcal{T}$ is Hausdorff, so that a and b are not separated by any pair of open sets. Then $(\forall U_a, U_b)(a \in U_a \wedge b \in U_b. \rightarrow .U_a \cap U_b \neq \emptyset)$. Then the collection of open sets that contain at least one of a and b has the finite intersection property and can be extended to an ultrafilter $\mathcal{U}$. But now a and b are distinct convergence points for $\mathcal{U}$. Notice that we have used the prime ideal theorem. ■

Remark 4 *A space is compact iff every ultrafilter on it has at least one convergence point.*

Proof:

$L \rightarrow R$

Suppose $\mathcal{T}$ is compact and let $\mathcal{U}$ be an ultrafilter on $\mathcal{T}$. We want $\mathcal{U}$ to have a convergence point. Consider the collection of closed subsets of $\mathcal{T}$ that are in $\mathcal{U}$. Is the intersection of them all empty? No, because if it were empty then (by compactness) an intersection of finitely many of them would be empty, contradicting the fact that $\mathcal{U}$ has the finite intersection property. So the intersection of all those closed sets is a nonempty closed set, so let x be a member of the intersection. Let O_x be an open neighborhood of x. We want $O_x \in \mathcal{U}$. If it isn't then its complement is a closed set in $\mathcal{U}$. But then x does not belong to the intersection of all closed sets in $\mathcal{U}$.

$R \to L$

For the other direction we use the definition of compactness according to which a space is compact if every set of closed sets with the fip has nonempty intersection. Let $\mathcal{T}$ be a space, and let $\mathcal{X}$ be a family of closed subsets with the fip. We must show $\bigcap \mathcal{X}$ is nonempty. By assumption $\mathcal{X}$ has fip, so – by BPI – $\mathcal{X}$ can be extended to an ultrafilter $\mathcal{U}$. By assumption $\mathcal{U}$ has a convergence point x. Every open neighborhood U_x of x is in $\mathcal{U}$. If $X \in \mathcal{X}$ then $X \in \mathcal{U}$; $U_x \in \mathcal{U}$, so $X \cap U_x \in \mathcal{U}$. We want $x \in X \cap U_x$. If it isn't, then $U_x \setminus X$ is an open set containing x that is not in $\mathcal{U}$ (since it is disjoint from X, which is in $\mathcal{U}$) contradicting the assumption that x was a convergence point of $\mathcal{U}$.

■

COROLLARY 2 *A nonempty space is compact Hausdorff iff every ultrafilter on it has precisely one convergence point.*

Products

We need to know how to relate ultrafilters on the product to ultrafilters on the factors.

Let $\{\mathcal{T}_i : i \in I\}$ be a family of topological spaces, and suppose that, for each $i \in I$, $\mathcal{U}_i$ is an ultrafilter on $\mathcal{T}_i$. How are we to define an ultraproduct $\mathcal{U}$ on the product $\prod_{i \in I} \mathcal{T}_i$? Let us say a subset X of the product is **large** if $(\forall i \in I)(\{f(i) : f \in X\} \in \mathcal{U}_i)$. So $\mathcal{U}$ is to be the set of all large subsets of $\prod_{i \in I} \mathcal{T}_i$? No reason to suppose that is an ultrafilter, but at least it can *generate* an ultrafilter.

For the other direction. Take an ultrafilter $\mathcal{U}$ on the product; its projections $\mathcal{U}_i$ to the factors are ultrafilters (on the factors) as follows.

$$\mathcal{U}_i = \{X \subseteq \mathcal{T}_i : \{f \in \prod_{i \in I} \mathcal{T}_i : f(i) \in X\} \in \mathcal{U}\}$$

$\mathcal{U}_i$ is an ultrafilter on $\mathcal{T}_i$.

It might be worth verifying my hunch that altho' we can get $\{\mathcal{U}_i : i \in I\}$ from $\mathcal{U}$, we cannot recover $\mathcal{U}$ from $\{\mathcal{U}_i : i \in I\}$

Next we prove

THEOREM 3 *Tikhonov's theorem for Compact Hausdorff spaces is equivalent to the Prime Ideal Theorem.*

Proof:

$R \to L$

So suppose we want to show that an arbitrary product of compact Hausdorff spaces is compact Hausdorff. Take your family $\{\mathcal{T}_i : i \in I\}$ of compact Hausdorff

spaces. We want to show that every ultrafilter on the product space $\prod_{i\in I}\{\mathcal{T}_i : i \in I\}$ has precisely one convergence point.

If the product is empty it is compact Hausdorff (check the small print) so let's assume that the product is nonempty.

So we have to show that every ultrafilter on the product space has precisely one convergence point. Take an ultrafilter $\mathcal{U}$ on the product; its projections to the factors are ultrafilters $\mathcal{U}_i$ (on the factors) as follows, as we saw above:

$$\mathcal{U}_i = \{X \subseteq \mathcal{T}_i : \{f \in \prod_{i\in I} \mathcal{T}_i : f(i) \in X\} \in \mathcal{U}\}$$

$\mathcal{U}_i$ is an ultrafilter on $\mathcal{T}_i$. By assumption on $\mathcal{T}_i$ (that it is compact Hausdorff) $\mathcal{U}_i$ has precisely one convergence point – call it "$k(i)$". The function k in the product that picks up these convergence points is the obvious suspect for a convergence point for $\mathcal{U}$. What is an open neighborhood of k? Let $I' \subseteq I$ be finite. Then $\{f \in \prod_{i\in I} \mathcal{T}_i : (\forall i \in I')(f(i) = k(i))\}$ is a neighborhood of k. Call it $O_{I'}$. By definition of k, we have $O_{I'} = \{f \in \prod_{i\in I} \mathcal{T}_i : (\forall i \in I')(f(i) = k(i))\}$. We want $O_{I'} \in \mathcal{U}$. $k(i)$ is a unique thing s.t. every open ball round it is in $\mathcal{U}_i$, so we can rewrite our definition of $O_{I'}$ by replacing '$f(i) = k(i)$' by '$(\forall X \in \mathcal{O}(\mathcal{T}_i))(f(i) \in X \to X \in \mathcal{U}_i)$' obtaining

$$O_{I'} = \{f \in \prod_{i\in I} \mathcal{T}_i : (\forall i \in I')(\forall X \in \mathcal{O}(\mathcal{T}_i))(f(i) \in X \to X \in \mathcal{U}_i)\}$$

Next (using the definition of $\mathcal{U}_i$ above) we can unwrap '$X \in \mathcal{U}_i$' as $X \subseteq \mathcal{T}_i \wedge (\{g \in \prod_{i\in I} \mathcal{T}_i : g(i) \in X\} \in \mathcal{U})$ and substitute to obtain

$$O_{I'} = \{f : (\forall i \in I')(\forall X \in \mathcal{O}(\mathcal{T}_i))((f(i) \in X \to \{g : g(i) \in X\} \in \mathcal{U})\}$$

. . . and i cannot for the life of me see why this object has to be in $\mathcal{U}$. . .

L $\to$ R

For the other direction we have to show how to deduce the Prime Ideal Theorem from the assumption that a product of compact Hausdorff spaces is compact Hausdorff. This is a classic exercise in pattern-matching.

Do we need the word 'nonempty' in here somewhere?

Chapter 7

tf teaches himself some Groups, Rings and Modules with a little help from his friends

7.1 A Handout for Steve Pike on Rings and Ideals

You know about $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{C}$. All these number systems have addition, and multiplication, and a 0 and a 1. The 0 is an additive unit and the 1 is a multiplicative unit. Some of them have additive and multiplicative *inverses*. 0 is not only an additive unit but a multiplicative *annihilator* (I love that word!). That is to say that $0 \cdot x = 0$ (in contrast to $1 \cdot x = x$, 1 being the multiplicative *unit*. You're probably wondering if there is an *additive* annihilator, which would be a bad thing ☠ such that $x + ☠ = ☠$; but there is no such thing, so you don't need to worry about it[1].)

These number systems are probably the right point of departure, the right place to launch yourself off from. The place you launch yourself *towards* contains lots of weird novel structures with scary names, and it would be nice to know what they mean and what they do, and why we should want to know.

A **ring** is a thing with multiplication (often written $\cdot$), addition (usually written $+$), a multiplicative unit (usually written '1'), an additive unit (which is also a multiplicative annihilator) usually written '0'. Those are the – so to speak – *ingredients* of the ring. The *recipe* for rings contains instructions like

[1]The symbol '∞' looks as if it might denote an additive annihilator: think of '$\infty + 1 = \infty$' but this is a false friend: ignore it.

commutativity of multiplication and of addition and distributivity of $\cdot$ over $+$: $x \cdot (y + z) = x \cdot y + x \cdot z$.

We start off with $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{C}$. These are – all of them – things you would call *numbers*. $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{C}$ (but not $\mathbb{N}$) are rings, and their shared behaviour of $\cdot$, $+$, 0 and 1 is what motivated the invention of the idea of a ring. God knows why they're called 'ring' (It's the same in French – *anneau*). There are other kinds of numbers too, that you haven't yet met, but will. I don't just mean integers-mod-p but things called *ordinals*[2] which are a kind of transfinite generalisation of $\mathbb{N}$, the quaternions (which are related to the complexes) and p-adics which you almost certainly won't need (but they're fun). As i say, all these things are numbers, in that they all seem to deal with, formalise etc, *quantity* of some kind. However, the gadgetry of numbers is quite *multifarous* (multiplication, addition, 0 and 1, inverses ...) and it's quite a good idea to build it all up piecemeal.

One thing to bear in mind: it is important to think of all these number systems as distinct entities, rather than as parts of a whole. The thing we think of as *the* number 1, for example, is actually a cluster of lots of entirely different things, not one (ha!) thing. The natural number 1 (the cardinal number 1) is the quantum of addition-and-subtraction of discrete multitudes (It's the smallest increment or decrement one can have) The real number 1 is the unit of multiplication of real-valued quantities such as length. There are stories to be told about the integer 1, the rational 1 and the complex 1; they're all different. It's important to think they are all different beco's if you try to think of them as all being the same thing then you will find it harder to reason about the 0s and 1s in the rings etc that are soon to come your way. You will be a victim of an *error of attachment*. The multiplicative unit of a ring you picked up in the gutter is just that – a multiplicative unit of a ring: it's not the same as the rational number 1, and you will waste time and effort if you try to think of it as the rational number 1. As Bishop Butler said: *Everything is what it is and is not another thing.*

There are two stories one can tell about number systems. One is that each one arises by abstracting from some particular set of natural phenomena, each is an attempt to give an abstract treatment of some naturally occuring concept of quantity. I think this is a good story, but i have professional colleagues who think it is fanciful.

The other story is that each kind of number arises from earlier kinds by adding inverses for certain operations. We start with $\mathbb{N}$, which has addition but does not have additive inverse; so we invent additive inverses and that gives us $\mathbb{Z}$. $\mathbb{Z}$ has multiplication but does not have multiplicative inverses, so we invent them and get $\mathbb{Q}$. $\mathbb{Q}$ has holes so we fill in the holes and get $\mathbb{R}$. (Don't worry if you don't know how to do that yet)

This is connected to the business of adding new objects to a domain to have inverses or (more generally) solutions to equations. Hence these stories about how to construct $\mathbb{Z}$ from $\mathbb{N}$, and so on. This is philosophically and technically

[2] ordinals aren't a ring so perhaps forget that i mentioned them

charged. For example: people lost a lot of sleep (as you may know) over the invention of complex numbers.

How do we get rings? Don't worry for the moment about what the ring axioms are; the point is that there are various structures out that that have a lot in common, and we abstráct that-which-they-have-in-common and call that ábstract thing a 'ring'. The obvious example of a ring that isn't a ring of numbers is the ring of (positive and negative) integers under addition. There are some other nice examples – the ring of polynomials with integer coefficients for example. You can add two such polynomials in the obvious way, and there is an obvious additive inverse of any poly – just swap plus and minus signs. You can multiply two such polynomials in the obvious way but one-over-a-polynomial is not another polynomial. (It might be a power series but that doesn't count).

7.1.1 Associativity, Commutativity and Distributivity

There are three notions – properties of operations – which have names. These ideas may be novel, and we should say something about them and where they come from. They are *Commutativity, Associativity and Distributivity.*

Steve: do i need to define these notions? Or do you know them already?

Commutativity

Addition tends to be commutative. You pile your goods into your van and drive off to market and sell them – in no particular order – out of the back of the van to people who come and see you. You are adding and subtracting stuff from the van contents. The *order* in which you pile goods into the van, and the order in which people buy them, alike have no effect on the amount you grow or sell or have in the van at the end of the day.

> Not all additions are commutative: ordinal addition isn't commutative but we won't worry about ordinals for the moment. In any case ordinals do not constitute a ring. Continue to forget that i even so much as mentioned them.

Associativity

Operations that correspond to actions have a tendency to be associative. My knitting patterns have instructions like "*knit a row of plain*"; "*knit two rows of purl*" The operation of concatenation on commands of this kind is clearly associative. So why is ring multiplication asociative? It's because – as we will see below – each ring element corresponds to an injection from the ring into itself, and composition of functions is associative, just like the concatenation on my knitting instructions is associative.

We need eventually to say something about why multiplication of octonions is not associative

Distributivity

Beginners can be a bit taken aback by distributivity and may wonder where it comes from. It's a bit of a mouthful, in that it's not (unlike associativity or

commutativity) a fact about a single operation in isolation, but a fact about how two operations interact.

If we have two binary operations o_1 and o_2 where o_1 distributes over o_2:

$$(\forall xyz)(o_1(x, o_2(y, z)) = o_2(o_1(x, y), o_1(x, z)))$$

(for example: $(\forall xyz)(x \cdot (y + z) = x \cdot y + x \cdot z)$)

what we are saying is that, for any x, the operation $y \mapsto o_1(x, y)$ is an endomorphism of the o_2 structure. Obvious when you think of it. On the integers, multiplication by a fixed integer is an endomorphism of the additive structure of the integers[3].

Do i need to explain *endomorphism?*

7.1.2 Whence cometh the idea of *ideal*?

Spoiler alert: there is a tone poem by Liszt called *Die Ideale*. Sadly it's nothing to do with our business here.

Keyword for this section is *Unique Factorisation*. This is something you will find reassuringly cuddly and familar. Think about $\mathbb{N}$, the set of whole numbers. You know about prime numbers. You also know (tho' you probably haven't given it much thought) that every whole number can be expressed as a product of primes, and that this product is unique "up to order" as they say. For example $6 = 2 \cdot 3$ so it's a product of two primes (2 and 3 are both primes) and multiplication is commutative so it doesn't matter which way round you reckon it. And – and this is the bit that is so obvious that you've never thought about it – this representation of 6 as a product of primes in *unique*: it's not going to turn out to also be 7×17 or anything like that.

I want you to think of unique factorisation as something rather special that might not happen in any-old ring, and the fact that it happens in the ring[4] $\mathbb{N}$ is a rather special fact. But to set it in a general ring-context we need to generalise the idea of a *prime* to a general setting. For this we need the notion of an *irreducible*. To get *irreducible* we need *invertible*.

- We first say an element of a ring is *invertible* iff it has a multiplicative inverse[5].
- Then we say that a non-zero non-invertible element is *irreducible* if it is not a product of two non-invertible elements.

[3] Digression: What i am now wondering is whether this is the correct insight to bring to bear on the challenge of understanding why dishonesty of one of $\vee$ or $\wedge$ means that the lattice fails to be distributive.

And what about the infinitary distributive law for Heyting algebras? $A \wedge (\exists x)B \longleftrightarrow (\exists x)(A \wedge B)$. This says that, for each open set, the operation of intersecting-with-it is a . . . homomorphism onto the subspace. . . ?

[4] Actually $\mathbb{N}$ isn't a ring since it doesn't have additive inverses but that doesn't matter. We still have factorisation and unique factorisation.

[5] Some people call an object with a multiplicative inverse a *unit*. I think they do it to annoy.

It might be an idea to spell out how this idea plays out in $\mathbb{N}$. In $\mathbb{N}$ no element (other than the unit 1) is invertible. So what are you doing if you are not a product of two noninvertible elements? You're a prime!

So in a general ring we have a notion of invertible/nonivertible element, so we have a notion of irreducible. So we can ask whether an ring element plucked at random can be represented as a product of irreducible elements, and if this representation is unique (up to order of multiplcation. If we and the ring are lucky – like $\mathbb{N}$ is lucky – then the answer is *Yes!* and we say that the ring **has unique factorisation**.

You can probably well believe, Dear Reader, that if a ring has unique factorisation then it is much easier to reason about it. There is a story that when Fermat wrote in the margin of a book that "I have a marvellous proof of this but the margin is too small to contain it" (he was talking sbout what we now call *Fermat's Last Theorem* – actually a theorem of Andrew Wiles) he did indeed have something like a proof, but it was based on the assumption that a particular ring had unique factorisation when, in actual fact, it didn't.

You can probably also believe, Dear Reader, that if you are studying a ring that annoyingly doesn't have unique factorisation, it might help if you expand that ring to a larger ring that *does* have unique factorisation so you can do your reasoning – whatever it was that you were trying to do – in that new ring instead.

So how do we get from a ring that doesn't have unique factorisation to a new, larger, ring that does? You add stuff. If you free-associate from this suggestion to the phenomenon of adding irrationals to $\mathbb{Q}$ (the rationals) so that $x^2 - 1 = 0$ has a solution then you will definitely be on the right track.

Now this project of *just adding stuff* is *prima facie* problematic. How do you know that the stuff you're adding is consistent? Philosophers have worried about this sort of move for years, as well they might. A famous example is *The Round Square*. It's no good postulating a round square, there ain't no such animal. The project of adding irrationals to $\mathbb{Q}$ had a much happier ending. This is because we *found a way of conceptualising the new objects*. There is a large literature on this, as you can probably imagine, but i want to skirt it to the extent that i can; nevertheless it's probably help to at least acknowledge that there is something there to worry about.

Could talk here about how we concretise integers, rationals starting from $\mathbb{N}$

A Standard Example

Let us navigate ourselves by first principles through the maze from a classic case of a ring that doesn't have unique factorisation to a ring that does.

Here is the standard example: $\mathbb{Z}[\sqrt{-5}]$ is sold to us as the substructure of $\mathbb{C}$ generated by $\mathbb{Z}$ and $\sqrt{-5}$.

> Digression **for Logicians only!**
> Now $\mathbb{C}$ is a field and every substructure of a field is an integral domain; the theory of integral domains is universal so an arbitrary intersection of integral domains is an integral domain, so the intersection of all substructures of $\mathbb{C}$ that contain [all the integers and] $\sqrt{-5}$ is an integral domain. [This

sounds *sooo* much like a logician's explanation; how do other mathmos do it?] It's generally known as $\mathbb{Z}[\sqrt{-5}]$.

If we have got as far as $\mathbb{Z}[\sqrt{-5}]$ we are happy about complex numbers, so i am going to assume, Dear Reader, that you are happy with complex numbers.

In $\mathbb{Z}[\sqrt{-5}]$ we can factorise 6 as $2 \cdot 3$ and also as $(1+\sqrt{-5}) \cdot (1-\sqrt{-5})$. Unique factorisation will fail if these four guys are all irreducible, beco's then 6 has *two distinct* expressions as a product of irreducibles.

So let's check that 2, 3, $1+\sqrt{-5}$ and $1-\sqrt{-5}$ are all irreducible in $\mathbb{Z}[\sqrt{-5}]$.
Suppose

$$1+\sqrt{-5} = (a+b\sqrt{-5})(c+d\sqrt{-5})$$

We multiply out the RHS to get

$$(ac+5bd)+\sqrt{-5}(bc+da)$$

This gives $ac+5bd=1$ and $bc+da=1$, and then

$$1-\sqrt{-5} = (ac+5bd)-\sqrt{-5}(bc+da)$$

and the RHS rearranges to

$$(a-b\sqrt{-5})(c-d\sqrt{-5})$$

whence
$(1+\sqrt{-5})(1-\sqrt{-5}) = (a+b\sqrt{-5})(c+d\sqrt{-5})(a-b\sqrt{-5})(c-d\sqrt{-5})$
and
$-4 = (a^2-5b^2)(c^2-5d^2)$
Now, if $a^2-5b^2 = \pm 1$, then $(a+b\sqrt{-5})(a-b\sqrt{-5}) = \pm 1$, so it is has a multiplicative inverse.
The same holds for c^2-5d^2.
So If $1+\sqrt{-5}$ is reducible, you have $a^2-5b^2 = \pm 2$, $c^2-5d^2 = \mp 2$ and that is impossible mod 4.

To my shame i had to look this up – never having had to work thru' it myself (my first degree was in History of Music and Philosophy!). This proof i found on StackExchange `https://stackexchange.com/sites#` supplied by a minor deity by the name of `Exodd`. May (s)he live for ever.

We also need to check that both 3 and 2 are irreducible in $\mathbb{Z}[\sqrt{-5}]$, but i think i can safely leave that to the reader.

So we invent "lower" factors – four of them in fact.

r_1 to be a common factor of 2 and $1+\sqrt{-5}$;
r_2 to be a common factor of 2 and $1-\sqrt{-5}$;
r_3 to be a common factor of 3 and $1+\sqrt{-5}$; and, finally
r_4 to be a common factor of 3 and $1-\sqrt{-5}$.

How are we to concretise these ficticious factors? Put entirely out of your head the pipe-dream of finding them in $\mathbb{C}$. They might be in $\mathbb{C}$ but the idea is to find a way of concretising them without making that assumption. After all we might want to add ideal divisors to rings that aren't rings of the form $\mathbb{Z}[\alpha]$ for a complex number α.

Remember here Quine's *bon mot* "no entity without identity". If we are to reason about these things at all we need to be able to tell them apart, and to recognise them when we see them. The key observation is that, although we (think) we do not know what these new roots are, we do at least know exactly what their nontrivial multiples are, and that gives us a way in. *We can tell them apart beco's they divide different things*. The map that takes elements-or-ideal-elements and sends each to the set of things it divides is injective; so how about we identify each element with the set of things it divides? (There is an obvious *prima facie* circularity problem here but we will ignore it in the hope that it will all come out in the wash.)

Let's consider the ideal element r_3 that divides 3 and $1+\sqrt{-5}$. Being a divisor of both 3 and $1+\sqrt{-5}$ is just the same (we reason in $\mathbb{Z}[\sqrt{-5}]$) as being a divisor of every element of $\mathbb{Z}[\sqrt{-5}]$ that is of the form $a\cdot 3 + b\cdot(1+\sqrt{-5})$ *and a divisor of nothing else*. We'd better spell this out. If x divides every element of $\mathbb{Z}[\sqrt{-5}]$ that is of the form $a\cdot 3 + b\cdot(1+\sqrt{-5})$ then by setting a to 1 and b to 0 we infer x divides 3 and by setting a to 0 and b to 1 we infer x divides $1+\sqrt{-5}$.

For the other direction if x divides both 3 and $1+\sqrt{-5}$ then x divides every element of $\mathbb{Z}[\sqrt{-5}]$ that is of the form $a\cdot 3 + b\cdot(1+\sqrt{-5})$.

Different ideal divisors will correspond to different subsets of $\mathbb{Z}[\sqrt{-5}]$, so we concretise r_3 *as that set*: $\{a\cdot 3 + b\cdot(1+\sqrt{-5}) : a,b\in\mathbb{Z}\}$.

Notice that we can tell r_1, r_2 ...apart in this way just by reference to other members of $\mathbb{Z}[\sqrt{-5}]$ ...we don't need to examine their relations to novel members of the new ring.

> A historical note: this response "think of them as sets!" to this concretisation challenge was one of the things that led to the birth of Set Theory.

> It might help calm the nerves to see what happens if we add to $\mathbb{Z}$ an ideal divisor of 81 and 48 by this method. The set we get is the set of all numbers of the form $81x+48y$ for $x,y\in\mathbb{Z}$. This set is precisely the set of all integer multiples of 3, and 3 is of course the HCF of 81 and 48.

Notice that if we concretise ideal divisors as sets in this way then there is nothing to stop us thinking of other elements of the ring as the set of things they divide, and then we have a very neat account of multiplication of *all* elements, ideal elements and ordinary elements alike: it's just $\cap$ – intersection. Observe that these ideal-elements-concretised-as-sets are closed under multiplication by ring elements

Thus the ring that we obtain by adding ideal divisors to $\mathbb{Z}[\sqrt{-5}]$ is an object whose members are subsets of $\mathbb{Z}[\sqrt{-5}]$. For each number $\alpha\in\mathbb{Z}[\sqrt{-5}]$ it contains the set $\{\alpha\cdot x : x\in\mathbb{Z}[\sqrt{-5}]\}$, and it also contains all the (sets-of-multiples

corresponding to the) ideal divisors. Every member of the new ring is a set, a subset of the old ring, and all members are there on the same terms, as it were. All of them are sets.

Now we have to explain what addition and multiplication are in this new ring.

Multiplication is easy: since we identify an element with the set of elements that it divides, then multiplication corresponds to intersection. Not so fast! It corresponds to the *closure-under-linear-combination* of the intersection. The intersection corresponds to the LCM not the HCF!

But what about *addition*?

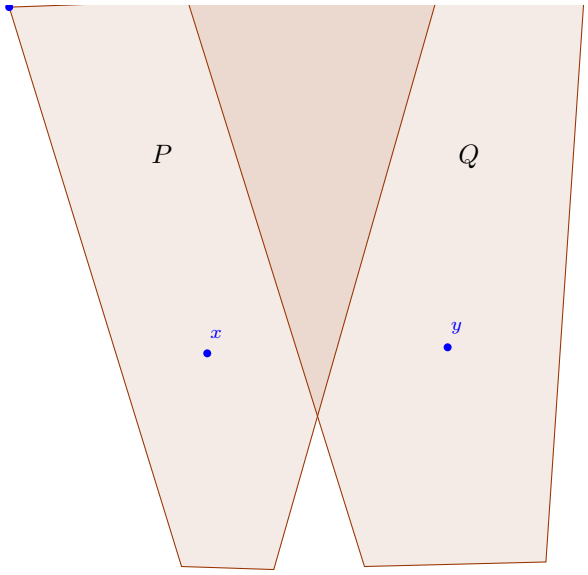
We take the set of all sums $x + y$ where $x \in P$ and $y \in Q$. Except we don't; we do it only when there is λ s.t. x is $\lambda \cdot a$ and y is $\lambda \cdot b$ (where a is the ideal generator of P and b is the ideal generator of Q). But how can we recover λ? We can't! Panic!! But it doesn't matter. What we can do, given $x \in P$ and $y \in Q$, is detect when they have the same λ. That happens when $x/a = y/b$ ... which of course doesn't make sense if P (or Q) is an ideal; however $x \cdot b = y \cdot a$, with a bit of effort, *does* make sense. It is equivalent to

$$\{w \in Q : x|w\} = \{z \in P : y|z\}$$

since the LHS is the set of multiples of x that are also multiples of b and the RHS is the set of multiples of y that are also multiples of a. So $P + Q$ is

$$\{x + y : x \in P \wedge y \in Q \wedge \{w \in Q : x|w\} = \{z \in P : y|z\}\}$$

We'd better check that this works!!!



We will want $P + Q$ to be an upper set under multiplication. So we want to know that if x and y are related by $\{w \in Q : x|w\} = \{w \in P : y|w\}$ then so are $x \cdot z$ and $y \cdot z$. Let's check it.

Now

$$\text{“}x \text{ and } y \text{ are related by } \{w \in Q : x|w\} = \{w \in P : y|w\}\text{”}$$

is equivalent to

$$(\forall w \in P \cap Q)(x|w \longleftrightarrow y|w)$$

and we want this to imply, for an arbitrary z,

$$(\forall w \in P \cap Q)((x \cdot z)|w \longleftrightarrow (y \cdot z)|w)$$

But this is immediate, since if the biconditional in the first holds for all $w \in P \cap Q$ it certainly holds for all such w that are multiples of z.

It looks as if it might be easy to verify distributivity of multiplication over addition.

Let R be another upper set (ideal). Then $R \cap (P + Q)$ is

$$R \cap \{x + y : x \in P \wedge y \in Q \wedge \{w \in Q : x|w\} = \{w \in P : y|w\}\}$$

$$\{x + y \in R : x \in P \wedge y \in Q \wedge \{w \in Q : x|w\} = \{w \in P : y|w\}\}$$

So: if $z \in R \cap (P + Q)$, it is $x + y$ for $x + y \in R$ and x and y satisfying various conditions. We want that to be a neccessary and sufficient condition for z to be in both $R + P$ and $R + Q$.

If z is in both $R + P$ and $R + Q$ then it is both $x + y$ with $x \in R$ and $y \in P$ (plus conditions) and $x' + y'$ with $x' \in R$ and $y' \in Q$ (plus conditions).

Ooops – we don't mean $R \cap (P + Q)$, we mean the *closure under linear combinations of* $R \cap (P + Q)$

7.1.3 Rings as bundles of endomorphisms

Steve: you can probably skip this section

Anyway when we start looking at rings we will connect this last thought about distributivity with the idea that every ring is the set of endomorphisms of an abelian group. Multiplication by a ring element is the execution of a homomorphism, so *of course* it distributes over addition. And *of course* since each ring element represents an action – and multiplication of ring elements is concatenation of actions – we expect multiplication to be associative. (No reason to expect it to be commutative)

Notice, too that this fact (that the distribution of o_1 over o_2 means that o_1 on the left is an o_2-homomorphism) makes particularly good sense when the o_2 structure is algebraic and accordingly is preserved under homomorphism.

John Howe tells me that every ring-with-1 is isomorphic to a subring of the ring of homomorphisms of some abelian group into itself. (what is the addition and what is the multiplication? He says:

"One considers the underlying abelian group $\langle R, + \rangle$ of our ring R and for each element r of the ring, define a map by $x \mapsto rx$, check it's a hom. Then define a map from R to $\mathrm{End}(\langle R, + \rangle)$ taking r to the multiply-by-r map, check that's a ring hom, then it's injective because there cannot be non-zero r that multiplies by 1 to 0. It's one of the nicer ploddy checks because it involves all the ring axioms for R."

A ring-with-a-unit is called a **unital** ring. Why would one ever be interested in rings without units? A ring without a unit is sometimes called a rng.

Q: What do you call a fish with no eyes?
A: a fsh[6] .

But actually you don't have to worry too much about rings without multiplicative units. Or – more correctly – a ring in which $0 = 1$. There is only one such ring – or rng – as follows:

$$x = 1 \cdot x = 0 \cdot x = 0$$

But x was arbitrary, so the ring contains only one element, namely the monster that is both 1 and 0.

H I A T U S

Multiplication by a ring element x is a group homomorphism and the range is the ideal (x). (What a truly terrible notation!!) However not every ideal arises that way and not every group homomorphism is multiplication by a ring element.

The ideals in a ring do not naturally form a ring.

Add $\sqrt{2}$ to $\mathbb{Q}$ and generate a ring. Is the result a field? Yes, beco's the difference-of-two-squares will give us $(a + b\sqrt{2})^{-1}$ – multiply top and bottom by $(a - b\sqrt{2})$. In fact (I am reassured by Imre) the same works for any algebraic. Indeed there is even a converse! A real number α is algebraic iff the ring generated by α over $\mathbb{Q}$ is a field.

$\mathbb{Z}[x_1 \ldots x_n]$ is a universal object for the class of identities involving only $\mathbb{1}$, 0, $\times$ + with n variables.

An identity is true in $\mathbb{C}$ iff it is true in every commutative ring.

An Imre example sheet question: which abelian groups can be the multiplicative group of a commutative ring? Randall says, think not just about 1

[6]
What do you call a deer with no eyes?
No idea;
What do you call a deer with no eyes and no goolies?
No fucking idea;
What do you call a deer with no eyes and no goolies and no *legs?*
Still No fucking idea.

and 0 but also -1! Are -1 and $+1$ the same? If so, then the additive group is of exponent 2, and that narrows things down a bit. If they are not then do the following. $0 = (-1) + 1$ so

$$\begin{aligned} 0 &= ((-1)+1)^2 \\ &= (-1)^2 + 1\cdot(-1) + 1\cdot(-1) + 1^2 \\ &= (-1)^2 + (-1) + (-1) + 1 \\ &= (-1)^2 + (-1) \\ 1 &= 1 + (-1)^2 + (-1) \\ 1 &= (-1)^2 \end{aligned}$$

So the multiplicative group contains an element of order 2.

So the idea now seems to me to be that we stumble into ring theory by starting to think about abelian groups and homomorphisms (some homomorphisms) from such groups to themselves. The honest way to do this would be to have a two-sorted or possibly even outright second-order theory to do this, but that would be *f-a-r* too sensible.

OK, so we have an abelian group R, and a semigroup of homomorphisms $R \to R$ containing at least the identity homomorphism and the zero homomorphism that sends everything to the additive unit (the 0) of the group. So we encode the homomorphisms somehow as elements of the group and encode their action by inventing another operation. This operation is called *multiplication* and is written with a '$\cdot$', so that $a \cdot x$ is that group element that x is sent to by the homomorphism encoded by the element a. This means that $\cdot$ has to distribute over $+$. However $a \cdot b$ is also the element of the group that encodes the composition of the two homomorphisms encoded by a and by b, and this second feature means that $\cdot$ has to be associative.

We always encode the identity homomorphism (that's what the $\mathbb{1}$ is for, and we identify an element of R to serve as the multiplicative unit) and we encode the annihilator (the zero homomorphism) with the additive unit of the abelian group, which we write as '0'.

However if we think of multiplication as arising from composition of homomorphisms there is no obvious reason why it should be commutative. And indeed, commutativity of multiplication is not part of the definition of a ring. Indeed there is a theorem of Wedderburn that says that a finite ring wherein every nonzero element has a multiplicative inverse is actually commutative. It would be nice to have an illuminating proof of this fact. There might be some very good reason for it to be true. The fact that the Unabomber had a nice proof of it cuts no ice. `https://en.wikipedia.org/wiki/Ted_Kaczynski`

It also seems to me that there is nothing in the ring structure to tell you which homomorphism marries up (*via* a bijection m) with which ring element – one could perhaps have permutation models of rings, obtained by composing m with a permutation of the semigroup of homomorphisms.

The last question on Imre's second sheet for GRM is "Can every abelian group be turned into a ring?"

0 and 1 in fields are the sum and product of the empty set of ring elements. The universal set of ring elements doesn't enter into it. But the `false` can also be the conjunction of the set of all formulæ. What are the sum and product of the whole carrier set? Jules sez: *prima facie* doesn't make sense unless the ring is finite. Try integers mod 8. You get 4. In fact you get the sum of the elements that are their own inverse.

[If the group is abelian we can define an addition operation on the homomorphisms by defining it pointwise.]

Can one say anything about these homomorphisms? Yes: distinct ring elements must encode distinct homomorphisms. Suppose two ring elements a and b encode the same homomorphism. Then $(\forall x)(a \cdot x = b \cdot x)$ so, in particular, we have $a \cdot \mathbb{1} = b \cdot \mathbb{1}$ whence (since we have decided that $\mathbb{1}$ encodes the identity homomorphism) $a = b$. Whew! Elements that encode homomorphisms that are not surjective cannot have multiplicative inverses. If every homomorphism is surjective does that mean we have a field or at least a skew-field? I can't see it: just beco's the homomorphism has an inverse doesn't mean that that inverse is coded by a ring element; the bundle of homomorphisms that we are resolved to encode are merely a semigroup after all, not a group.

If all the (nonzero?) homomorphisms are permutations then the ring is an integral domain. [prove this!] Presumably the converse is not true.

As we have seen, every ring element corresponds to a ring homomorphism (multiplication by that ring element). But not every ring homomorphism corresponds to an element; if you push it you can think of every homomorphism as an *ideal element*.

Multiplication by a ring element is a group homomorphism (but not a ring homomorphism) and the range of the homomorphism is an ideal.

If every homomorphism encoded by an element is injective then the ring is an integral domain. . . ?

What do we want to say about $0 \neq \mathbb{1}$?

We have an abelian group G, and a semigroup *Hom* of homomorphisms $G \to G$ containing the identity homomorphism $\mathbb{1}$ and the annihilator 0. There is a bijection $\mathfrak{k}{:}G \longleftrightarrow \mathit{Hom}$. We stipulate that $\mathfrak{k}(\mathbb{1}_G) = \mathbb{1}$ but in all other respects $\mathfrak{k}$ can do what it likes.

Notice that if the annihilator (the identically zero function) is a homomorphism then the singleton of 0 is a ring, and a ring in which $0 = 1$. So rings are not obliged to be unital So if we *both* want annihilators *and* want to think of all elements as homomorphisms *then* we need to hang onto the possibility of rings not being unital.

> The point is not that non-unital rings are particularly useful; the point is that we need to admit them in order to make the general theory smoother. Specifically (as the above shows) if we want a homomorphic image of a ring to be a ring.

Now comes the trickery. We introduce a new binary function symbol '$\cdot$' and require that '$a \cdot x$' be syntactic sugar for '$\mathfrak{k}(a)(x)$'. However we also want

'$\mathfrak{k}(a \cdot x)$' to denote the composition $\mathfrak{k}(a) \circ \mathfrak{k}(x)$ ($\circ$ is composition of functions.) Now composition of functions is associative, so we we must have both

$$a \cdot x \cdot y \;=\; (a \cdot x) \cdot y \;=\; (\mathfrak{k}(a)(x)) \cdot y \;=\; \mathfrak{k}((\mathfrak{k}(a)(x)))(y)$$

and

$$a \cdot x \cdot y \;=\; a \cdot (x \cdot y) \;=\; (\mathfrak{k}(a))(\mathfrak{k}(x)(y)).$$

whence

$$\mathfrak{k}((\mathfrak{k}(a)(x)))(y) = (\mathfrak{k}(a))(\mathfrak{k}(x)(y)).$$

which should give us a nice commutative diagram. picture here

The thing that is puzzling me is: if we are given Hom, a semigroup of homomorphisms $G \to G$ containing the identically zero function (the annihilator 0) and the identity map $\mathbb{1}$, with the information that $|Hom| = |G|$, how are we to find $\mathfrak{k}$ having these special properties?

We do the following. Let $\mathfrak{k}$ be *any* bijection $G \to Hom$ sending the 0 of the group to the annihilator, and we define $\cdot$ in terms of it, thus:

$$a \cdot b = \mathfrak{k}^{-1}(\mathfrak{k}(a) \circ \mathfrak{k}(b)). \tag{1}$$

But we also want $a \cdot b = \mathfrak{k}(a)(b)$!!

This gives

$$\mathfrak{k}(\mathfrak{k}(a)(b)) = \mathfrak{k}(a) \circ \mathfrak{k}(b). \tag{2}$$

How do we swing that??

Perhaps we don't .... Perhaps we say the following. Suppose Hom is a semigroup of homomorphisms $G \to G$ containing the annihilator and the unit such that there is a bijection $\mathfrak{k} : G \to Hom$, satisfying (2), then we define $\cdot$ as in (1).

Let's start with what we know from first year algebra. We know about $\mathbb{N}$, about $\mathbb{Q}$, about $\mathbb{Z}$ and $\mathbb{C}$. We know what a field is (it has $+$ and $\cdot$ and $-$ and 0 and $\mathbb{1}$ and multiplicative inverses). So what is $[\langle a, b\rangle]_\sim + [\langle x, y\rangle]_\sim$? Can it be anything as simple as $[\langle a + x, b + y\rangle]_\sim$? That looks promising. So what is the injection from $\mathbb{N}$ into the quotient $\{[\langle a, b\rangle]_\sim : a, b \in \mathbb{N}\}$? I think it must be the map $a \mapsto [\langle a, 0\rangle]_\sim$ – which is at least injective. If i was right about addition then $\langle a, a\rangle \sim \langle 0, 0\rangle$ so the additive inverse of $[\langle a, 0\rangle]_\sim$ presumably must be $[\langle 0, a\rangle]_\sim$.

So we use the same trick to add multiplicative inverses to a ring, as it might be, $\mathbb{Z}$. But now we have to think about how we define addition on the $\sim$-equivalence classes. Analogy with $\mathbb{Q}$ and $\mathbb{Z}$ suggests that the addition operation on pairs $\langle a, b\rangle$ and $\langle x, y\rangle$ must be $\langle a \cdot y + b \cdot x,\ b \cdot y\rangle$. However we really do have to check that $\sim$ really is a congruence relation for this operation. Mind you, this is probably all in some Wikipædia article on algebra.

7.1.4 Ideals and Homomorphisms

For the moment – until we nail all that down – we have to think of ideals as ranges of homomorphisms. So suppose we have two homomorphisms h_1 and h_2. Then $h_1 + h_2$ of x is obviously going to be $h_1(x) + h_2(x)$. Check that this really is a homomorphism. We want

$$(\forall xy)((h_1 + h_2)(x + y) = (h_1 + h_2)(x) + (h_1 + h_2)(y)).$$

The RHS is

$$h_1(x) + h_2(x) + h_1(y) + h_2(y)$$

which we want to be equal to

$$h_1(x) + h_1(y) + h_2(x) + h_2(y)$$

This is where we discover that + has to be abelian!

Incidentally, are these ideals prime? Wikipædia sez that an ideal P in a ring R is prime *if [whenever] a and b are two elements of R such that their product $a \cdot b$ is an element of P, then $a \in P \vee b \in P$.* (Think: *prime number*: if p is a prime number then the set of multiples of p is a prime ideal in $\mathbb{Z}$.)

Let's consider again the ideal element – call it r_1 – that divides 3 and $1+\sqrt{-5}$. We ask: is it the case that whenever $x \cdot y$ belongs to this ideal (and x and y are both in $\mathbb{Z}[\sqrt{-5}]$) then at least one of x and y does? The ideal consists of precisely those things in $\mathbb{Z}[\sqrt{-5}]$ that are divisible by 3 or by $1 + \sqrt{-5}$. (Or both). Such a thing is a complex number of the form $3 \cdot a + (1 + \sqrt{-5}) \cdot b$ with $a, b \in \mathbb{Z}$. More usefully, it is $(3a + b) + b \cdot \sqrt{-5}$.

Now suppose

$$(3a + b) + b \cdot \sqrt{-5}) \;=\; (c + d\sqrt{-5})(e + f\sqrt{-5}) \tag{1}$$

Why are there two 'b's on the LHS?

We want to show that one of the factors on the RHS is of the form $(3x + y) + y \cdot \sqrt{-5}$ with $x, y \in \mathbb{Z}$.

Multiply out the RHS of (1) to get

$$ce - 5df + (cf + ed)\sqrt{-5} \;=\; (3a + b) + b \cdot \sqrt{-5}$$

(We can perform this computation in $\mathbb{C}$) and we can identify real and imaginary parts getting

$$3a + b = ce - 5df \quad \text{and} \quad cf + ed = b$$

whence

$$3a = ce - 5df - cf - ed$$

so at the very least

$$3 \mid (ce - 5df - cf - ed).$$

Now $ce - 5df - cf - ed = e \cdot (c - d) - f \cdot (c + 5d)$. We exploit the fact that $c + 5d = (c - d) + 6d$ to get

$$3 \mid (e - f) \cdot (c - d) + 6df$$

But $3 \mid 6df$ so this is equivalent to

$$3 \mid (e - f) \cdot (c - d)$$

whence

$$3 \mid (e - f) \vee 3 \mid (c - d)$$

.

The first horn corresponds to $c + d\sqrt{-5}$ being of the form $(3x + y) + y \cdot \sqrt{-5}$ with $x, y \in \mathbb{Z}$ and the second horn corresponds to $e + f\sqrt{-5}$ being of the form $(3x + y) + y \cdot \sqrt{-5}$ with $x, y \in \mathbb{Z}$.

There are three other ideals to consider in this case, and of course a whole general result to be proved but for the moment i am going to assume that this positive result is typical, even tho' the proof does seem frightfully *ad hoc*. ■

Observe that if the ideal element is represented as the set of things it's supposed to divide, then every genuine element, too, can be represented in the same way – as the set of things it divides. But of course this set is exactly the image of the ring in the homomorphism coded by that element!

Reflect, too, that multiplication of ideal-elements-tho'rt-of-as-sets is obviously $\cap$, which of course is the multiplication of boolean rings.

I'm having difficulty connecting these two ideas ...

(i) ideals as ideal elements (which seems to be the same as ideals-as-the-range-of-a-homomorphism) and

(ii) ideals as kernels of homomorphisms. Isn't there the concept of an e x a c t s e q u e n c e where the range of one homomorphism is the kernel of the next...?

OK, every ring element corresponds to a homomorphism $G \to G$. How do ideals correspond to homomorphisms $G \to G$? You can't recover a homomorphism from its kernel, since all automorphisms have the same kernel. You can recover the quotient, but that's not the same.

No: a ring element corresponds to a *group* homomorphism – multiply by that element. The range of the homomorphism is an ideal, which is the kernel of a *ring* homomorphism. But it's a different homomorphism.

The family of ideals in a ring has multiplicative and additive structure but it is not a nice ring – it certainly isn't the kind of thing you can embed your original ring in. For one thing, the additive group has index 2.

Wikipædia says:

> Commutative Rings $\supseteq$ Integral Domains $\supseteq$ Integrally Closed Domains $\supseteq$ Unique Factorization Domains $\supseteq$ Principal Ideal Domains $\supseteq$ Euclidean Domains $\supseteq$ Fields

Let's take these in order.

Commutative Rings

A ring is a set with $+$, making it an abelian group, and a multiplication that distributes over $+$. Clearly an algebraic theory. This fact is fundamental.

Every element corresponds to a homomorphism, but possibly not every homomorphism corresponds to an element. [Is this the condition for being a principal ideal domain??] Kernels of homomorphisms are *ideal elements* or just *ideals*. I think the kernels of those homomorphisms that correspond to genuine elements are *principal* ideals, but i'd better check. Alternative definition: an ideal is a [normal – but then all subgroups are normal beco's the group is abelian] subgroup that is closed under multiplication by arbitrary elements of the ring. Being an ideal is a closedness property so we have a notion of an ideal generated by a bundle of elements.

Beco's the concept of ideal arises from the idea of a kernel of a ring homomorphism we always have a notion of quotient over an ideal. We need to take account of the fact that a ring homomorphism is a more specific notion than a group homomorphism. The kernel of a group homomorphism won't have any multiplicative structure; the kernel of a ring homomorphism, well ...

How are we to take account of the multiplicative structure? If h is a ring homomorphism then the corresponding ideal must be $\{x : h(x) = 0\}$. Now suppose x is in the ideal, so that $h(x) = 0$, and let y be any ring element whatever. Then $h(x \cdot y) = h(x) \cdot h(y) = 0 \cdot h(y) = 0$, so $x \cdot y$ is in the ideal as well. So, if ideals are to capture the multiplicative structure they must be closed under multiplication by arbitrary ring elements.

Let's check that we now have a good notion of quotient over ideals thus construed. Writing "$-x$" for the additive inverse of x, we can express the obvious equivalence relation as $x \sim_I y \iff x + (-y) \in I$. This is clearly going to be a congruence relation for the group operation, for the usual reasons. Must check that it is a congruence for the ring multiplication as well.... So, given $x \sim_I x'$ and $y \sim_I y'$ we want to infer $x \cdot y \sim_I x' \cdot y'$:

supply a proof

Now consider the homomorphism corresponding to a ring element a. The corresponding ideal will be $\{x \in R : a \cdot x = 0\}$. This set is clearly closed under addition and under multiplication by arbitrary ring elements. Indeed it is the $\subseteq$-least set containing (containing what??) closed both under addition and under multiplication by arbitrary ring elements.

Yes, containing what? I think any x s.t. $a \cdot x = 0$ will do as a seed, co's one gets any other one by the two operations of addition and multiplication by arbitrary ring elements.

Consider the ideal (a) generated by the element $a \in R$. It's an ideal, so presumably it corresponds to a homomorphism. Which homomorphism? Does it correspond to an element?

There seems to be no connection between the homomorphism $x \mapsto a \cdot x$ and the ideal (a). Is this beco's we are not told the bijection between the group and the set of homomorphisms to be encoded?

An ideal is *prime* iff the quotient over it is an integral domain. Wikipædia

sez that an ideal P in a ring R is prime *if [whenever] a and b are two elements of R such that their product $a \cdot b$ is an element of P, then $a \in P \vee b \in P$.*

Wikipædia says: "An element p of a commutative ring R is said to be *prime* if it is not zero or a unit and whenever p divides $a \cdot b$ for some a and b in R, then p divides a or p divides b. Equivalently, an element p is prime if, and only if, the principal ideal (p) generated by p is a nonzero prime ideal."

Integral Domain

Wikipædia says: "An *Integral Domain* is a commutative ring with an identity with no zero-divisors. That is $ab = 0$ implies $a = 0 \vee b = 0$."

The language of fields does not contain a function symbol for multiplicative inverse – that function is not total. It has $+$ and $\cdot$, 0 and $\mathbb{1}$ and additive inverse $-$. It's a $\forall^*\exists$ theory not a $\forall^*$ theory, since it has the axiom $(\forall x)(x \neq 0 \rightarrow (\exists y)(x \cdot y = \mathbb{1}))$. Every substructure of a field (a structure for the language of fields) is an integral domain. So the theory of integral domains is the $\forall^*$-fragment of the theory of fields. Every integral domain can be canonically enlarged (by field-of-fractions) to a field, and this enlargement has a universal property.

Zachiri sez that the theory of integral domains is the universal fragment (also) of the theory of algebraically closed fields.

The characteristic axiom of integral domains is

$$(\forall ab)(a \cdot b = 0 \rightarrow a = 0 \vee b = 0).$$

This is not Horn, so we shouldn't expect a product of integral domains to be an integral domain. $\mathbb{Z}$-mod-2 $\times$ $\mathbb{Z}$-mod-3 is not an integral domain (sez Zachiri). Nor should we expect a homomorphic image of an integral domain to be an integral domain: Thdre is an obvious homomorphism $\mathbb{Z} \twoheadrightarrow \mathbb{Z}$ (mod 6).

Finite integral domains are fields. Let a be any nonzero element and consider $a, a^2, \ldots a^n \ldots$. At some point you must get back to a. So, for some n, $a^n = a$. So $a^{n-1} \cdot a = a$. But then $a^{n-1} = \mathbb{1}$ and $a^{n-2} \cdot a = a^{n-1} = \mathbb{1}$. So a^{n-2} was the multiplicative inverse.

Not sure that i believe that, actually

Quotients and fields-of-fractions

An integral domain is that kind of a ring such that the field of fractions of it is a field. That true? No proof of this fact here – and there should be.

A ring is a *local ring* iff it has a unique maximal ideal. This is first-order but presumably not Horn. A quotient of a local ring over this unique maximal ideal is a field.

Apparently if there is a unique maximal ideal it is the set of noninvertible elements, and if that set is an ideal it is the unique maximal one.

We'd better prove this.

x is noninvertible: $(\forall y)(x \cdot y \neq \mathbb{1})$.

Closed under addition and scalar multiplication.

$$(\forall xx')((\forall y')(x \cdot y' \neq \mathbb{1} \wedge x' \cdot y' \neq \mathbb{1}) \to (\forall y)((x + x') \cdot y \neq \mathbb{1})).$$

$(\forall x)(\forall z)((\forall y)(x \cdot y \neq 1) \to (\forall y)((x \cdot z) \cdot y \neq \mathbb{1}))$.
The first one becomes

$$(\forall xx'y)((\forall y')(x \cdot y' \neq \mathbb{1} \wedge x' \cdot y' \neq \mathbb{1}) \to (x + x') \cdot y \neq \mathbb{1}).$$

which doesn't look horn to me, or even universal. It does look universal-existential.

Example: the ring of germs of cts functions $\mathbb{R} \to \mathbb{R}$ is local. (See Wikipædia) So: what is the unique maximal ideal? And what is the quotient field?

Is it true that if R/p is a field then p is the unique maximal ideal in R and is the ideal of noninvertible elements?

A *Boolean ring* is a ring whose additive group is a group of exponent 2.

Integrally Closed Domain

Wikipædia says: "An *Integrally Closed Domain* A is an integral domain whose integral closure in its field of fractions is A itself". In other words, anything in the field of fractions over A that is the root of a poly with coefficients in A is already in A.

Not sure what this does for us. Must check whether or not it is first-order/algebraic.

Unique factorisation Domain

A *Unique Factorisation Domain*. Wikipædia says: "a unique factorization domain (UFD) is a commutative ring in which every non-zero non-unit element can be written as a product of prime elements (or irreducible elements), uniquely up to order and units, analogous to the fundamental theorem of arithmetic for the integers."

Not clear whether or not this is first-order.

From UFDs onwards, prime elements and irreducible elements are the same concept.

Principal Ideal Domain

A *Principal Ideal Domain* or PID is an integral domain in which all ideals are principal. A key point is that there is a division algorithm in a PID: any two elements a, b have a greatest common divisor, which may be obtained as a generator of the ideal (a, b) generated by a and b.

I can't see whether or not this is a first-order condition.

Integral Domain

An integral domain in which we can run Euclid's algorithm is a *Euclidean domain.*

Wikipædia says "A principal ideal domain is an integral domain in which every proper ideal can be generated by a single element."

Sounds second-order to me.

Field

A field is a commutative ring wherein every nonzero element has a multiplicative inverse, so that the nonzero elements are an abelian group.

An afterthought to be sorted out

Suppose i try to extend the naturals by adding a common factor to two numbers which we know to be coprime – 5 and 7, say. The ideal divisor is the set $\{5a + 7b \in \mathbb{N} : a, b \in \mathbb{Z}\}$ which is of course $\mathbb{N}$ itself. So it doesn't work. (*Of course* it doesn't work!) Is this because $\mathbb{N}$ already has unique factorisation? Or is a principal ideal domain? Or what?

It's just struck me (5/ii/21 in Hargreaves st) that the correct definition of an ideal is something closed under linear combinations!

What sort of representation theorems are there for rings?

When can a ring have a prime ideal that isn't maximal? Well, it can't be boolean for a start. What else? "Prime" ideal reminds us that there is a notion of multiplication of ideals, which is intersection (as in boolean rings!) and a prime ideal is not the intersection of two ideals!

Not every abelian group is the additive group of a ring.

Just struck me (3/ii/23) how remarkable it is that the concretisations of ideal divisors turn out to be the kernels of homomorphisms!!!

A ring homomorphism respects addition and multiplication, and sends $\mathbb{1}$ to $\mathbb{1}$. This implies that it sends 0 to 0, and preserves additive inverse. The kernel of a homomorphism f is the preimage $f^{-1}\text{“}\{0\}$ of $\{0\}$. The point is that once we know which things are equivalent to 0 we can ascertain when two elements in the domain are equivalent (belong to the same fibre of f) – we can recover the quotient from the fibre $f^{-1}\text{“}\{0\}$: $f(x) = f(y)$ iff $x + -y$ belongs to the kernel. That doesn't mean we can recover f itself from the fibre $f^{-1}\text{“}\{0\}$ because we could compose f on the left with an automorphism of the codomain and get the same answer.

There is more to being a ring homomorphism than being a homomorphism of the additive group! $f^{-1}\text{“}\{0\}$ has to be a (normal, but then all subgroups are normal) subgroup of the additive group, but that isn't enough, co's the multiplicative structure has to be respected too.

7.2 Quaternions, a theorem of Hurwitz

Dunedin, August 2018. Bruce has made me listen to a talk on youtube about quaternions. There is this theorem that i remember vaguely that the complexes, the quaternions and the octonions are the only three examples of a something-or-other. I still don't know what a something-or-other is, but i did learn one thing. Suppose we want to add another flavour of number (reals and complexes are the first two) to $\mathbb{C}$. So we want to add a new kind of unit – write it $\mathfrak{j}$ – and we want $\mathfrak{j}^2 = -1$ (not sure why, but there you go) and we want $\mathfrak{i}$, 1 and $\mathfrak{j}$ to be linearly independent. We also want every number in this new system to be a linear combination of $\mathfrak{i}$, $\mathfrak{j}$ and 1. So in particular we want

$$\mathfrak{i} \cdot \mathfrak{j} = a + b \cdot \mathfrak{i} + c \cdot \mathfrak{j}$$

multiply both sides by $\mathfrak{j}$

$$\mathfrak{i} \cdot \mathfrak{j}^2 = a \cdot \mathfrak{j} + b \cdot \mathfrak{i} \cdot \mathfrak{j} + c \cdot \mathfrak{j}^2$$

but $\mathfrak{j}^2 = -1$ so we can simplify to

$$-\mathfrak{i} = a \cdot \mathfrak{j} + b \cdot \mathfrak{i} \cdot \mathfrak{j} - c$$

but we started off with an equation for $\mathfrak{i} \cdot \mathfrak{j}$ which we can substitute in

$$-\mathfrak{i} = a \cdot \mathfrak{j} + b \cdot \mathfrak{i} \cdot \mathfrak{j} - c$$

to get

$$-\mathfrak{i} = a \cdot \mathfrak{j} + b \cdot (a + b \cdot \mathfrak{i} + c \cdot \mathfrak{j}) - c$$

but then $\mathfrak{i}$ and $\mathfrak{j}$ are not linearly independent!

Hermann sez:

$$\langle v_0, v_1, v_2, v_3 \rangle \cdot \langle w_0, w_1, w_2, w_3 \rangle =$$

coordinate 0 =

$$v_0 \cdot w_0 - v_1 \cdot w_1 - v_2 \cdot w_2 - v_3 \cdot w_3$$

coordinate 1 =

$$v_0 \cdot w_1 + w_0 \cdot v_1 + v_2 \cdot w_3 - v_3 \cdot w_2$$

coordinate 2 =

$$v_0 \cdot w_2 + w_0 \cdot v_2 + v_3 \cdot w_1 - v_1 \cdot w_3$$

coordinate 3 =

$$v_0 \cdot w_3 \; + \; w_0 \cdot v_3 \; + \; v_1 \cdot w_2 \; - \; v_2 \cdot w_1$$

It seems they are 2 $\times$2 matrices over the complexes

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad i = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}; \quad j = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \quad k = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

But there is also the vector space of 3×3 matrices over the complexes. Why isn't that another thing like the quaternions and the octonions? What is special about them?

No such thing as a module over the octonions: function composition is associative and that enforces associativity in the algebra.

7.3 Notes from James' Lectures in about 2001

A poly is **monic** if the coefficient of its term of highest degree is 1.

Is 'principal' the same as 'finitely generated' when it comes to ideals in rings? No: think of ring $\mathbb{Z}[X, Y]$ of polys in two variables over $\mathbb{Z}$. The ideal generated by the two polys X and Y is finitely generated but not principal.

Remember that 'irreducible' doesn't mean 'has no factors'. '$(x^2 + 2)^2$' is not irreducible over $\mathbb{R}$, beco's it has a factor, even tho' it has no roots in $\mathbb{R}$. [this has got garbled]

Wikipædia sez: If F is a field, a non-constant polynomial is **irreducible over** F if its coefficients belong to F and it cannot be factored into the product of two non-constant polynomials with coefficients in F.

How nice is $R[X]$, the ring of polys in one vbl over R, as a function of the niceness of R? The following classes are closed under the ring-of-polys construction: Integral domains, unique factorisation domains, Noetherian rings (thm of Hilbert's which James sez is quite hard – isn't that *Hilbert's basis theorem*, secretly a theorem in WQO theory?) but not principal ideal domains.

Obvious when you think of it (which i didn't)... the ring of polys in *two* vbls over a ring R is just the ring of polys in *one* variable over ... the ring of polys in *one* variable over the ring R! Duh! For all that it might be worth writing out a proof...

Here's how to extend a field. Find an irreducible poly P in the field. This poly generates an ideal in $F[X]$. Consider the quotient. This is going to be a field for a number of special reasons.

Let F be a field. Then $F[X]$ is always a PID. If a and b are in $F[X]$ and $b \neq 0$ there are unique q and r s.t. $a = b \cdot q + r$ where $\deg(r) < \deg(b)$.

(For this you really do need F to be a field: you need to be able to multiply the leading coefficient of the divisor by something to get the leading coefficient of the dividend so you need multiplicative inverses). Not only that but you

get Euclid's algorithm. This will be useful here, as in the integers, for getting multiplicative inverses.

For a start $F[X]$ is a PID. Let I be an ideal. We will show it is principal. Let m be a poly in I of minimal degree, and a any other poly in I. Try dividing m into a. What is the remainder? It must be of degree less than $\deg(m)$, which, since m was of minimal degree, must mean the remainder is 0, so that a was a multiple of m. So I is principal. (Not sure why this matters)

Two polys in $F[X]$ become identified in $F[X]/P$ if their difference is in the ideal, which is to say if their difference is a multiple of P. The difference between two distinct field elements is a nonzero field element and that is not a multiple of P. This is beco's if you multiply P by 0 you get 0, and if you multiply it by a nonzero element you get another nontrivial polynomial: there is no way of zero-ing the leading term. The effect of this is that the quotient (assuming it is a field, which it will be in fact) is a proper extension.

Notice also that no two field elements get identified in the quotient.

James sez: every element of the quotient contains a poly of degree at most P. This seems to be important. To establish this we have to, on being presented with a poly Q, come up with a poly Q' which is equivalent to Q and is of degree less than degree of P. If $\deg(Q) <\deg(P)$ we're home and hosed. If not we divide P into Q and look at the remainder, R. By the division algorithm, $\deg(R) <\deg(P)$, and the difference between R and Q is a multiple of P, so we're done.

Notice that two distinct polys of degree $< \deg(P)$ cannot be equivalent, as their difference is also of degree $< \deg(P)$ and so cannot be a multiple of P. So every equivalence class contains a poly of degree $< \deg(P)$ and all such polys belong to distinct equivalence classes. There are clearly $|F|^{deg(P)}$ such polys, so there are $|F|^{deg(P)}$ such equivalence classes.

Now all we have to do is find a multiplicative inverse for Q. This is where we have to use the fact that P is irreducible and $F[X]$ is a PID, or at least a Euclidean domain, one where Euclid's algorithm works. We use Euclid's algorithm on Q and P. This will give us polys q and p such that $P \cdot q - Q \cdot p = hcf(P, Q)$ but the RHS is 1 beco's P is irreducible. Well actually the output of Euclid will be a constant on the RHS (something that has a multiplicative inverse, that's the point) and then the poly p is the inverse we want.

(Notice that we really do need irreducibility, not simply having-no-factors beco's if P were $(x^2 + 2)^2$ the hcf might be $(x^2 + 2)$.)

If E is a subfield of F then F can be tho'rt of as a vector space over E. So the additive group of any (finite) field is simply the product of finitely many cyclic groups. For each $x \in F$, $\lambda y.x \cdot y$ is a linear map from F (tho'rt of as a vector space) into itself and so has a determinant and a trace, which are elements of E. Apparently this is useful.

tf to Edmund

Nothing urgent. i wanted to continue picking your brains about finite fields. How am i to think of the elements of the field of size 49, say? Presumably not

as ordered pairs of integers mod 7...

Edmund to tf

Well, in fact this is not such a bad way. Since the field has dimension 2 as a vector space over the field of order 7, you can pick coordinates and do just that. You obviously want one of the basis elements to be 1, and the other can be picked as convenient. Anyway, the vector-space stuff gives you the addition on the field. The other basis element would give you the multiplication. Since the base field has order 7, the group of units has size 6 and so not everything is a perfect square (the size of the group is divisible by 2). So you can pick something which doesn't have a square root. It will get square roots in the field of size 49 (uniqueness of that field), and you can pick one of them as the basis element and do multiplication and addition as you'd expect for expressions of the form $a + b\sqrt{n} : (n = 3, 5, 6)$

I'm not sure how much this helps in general, for example whether you can find sensible representations of finite fields as quotients of subfields of the complexes.

Why can't we just take the intersection of all fields extending $\mathbb{R}$ and containing the desired root? The theory of fields is not Horn and so it isn't obvious that the intersection is a field. But actually it is, so that's not a problem. Something that gives me cause for tho'rt is that in order to construct the field extension that way one has to think of the fields concretely. Is there a universal-algebra way of thinking of it? But a more serious immediate problem is that if we magick the field into existence that way we can't so easily prove that it's of degree n over $\mathbb{R}$ where n is the degree of the poly.

It's OK if you are doing it all inside $\mathbb{C}$, which already exists. Then you really can just take the intersection blah.

If we want to construct a field extension explicitly, why not do it syntactically? Take the set of all words in $+$ and $\times$, and reduce by the obvious equations. Well, perhaps that's what the usual construction is, dressed up to make it look more idiomatic from the point of view of field theory.

Note: any field can be tho'rt of as a vector space over any subfield. (check this) For instance – this is a perverse example – $\mathbb{R}$ is a field of degree $2^{\aleph_0}$ over $\mathbb{Q}$. I think this must be how it works. Consider the equivalence relation among reals of having rational difference. (I think we mean difference not ratio). Pick one representative from each class. Then every real can be expressed as a sum of rational multiples... no, not quite. It's related to the question of additive functions $\mathbb{R} \to \mathbb{R}$. These are precisely the linear maps of $\mathbb{R}$ tho'rt of as a vector space over $\mathbb{Q}$.

$\mathbb{C}$ is an extension of degree 2 over $\mathbb{R}$, co's it's a Vector Space of dim 2 over $\mathbb{R}$. In contrast $\mathbb{R}$ is not an extension of $\mathbb{Q}$ of any finite degree – any finite degree extension of $\mathbb{Q}$ must be ctbl.

THEOREM 4
Suppose E is a subfield of F which is a subfield of G. The degree of G over E is the product of the degree of G over F, (written '$[G : F]$') and $[F : E]$.

Proof:

Suppose $[F : E] = m$ and fix $f_1 \dots f_m$ a basis for F as a vector space over E.

Suppose $[G : F] = n$ and fix $g_1 \dots g_n$ a basis for G as a vector space over F.

It turns out that $\{f_i \cdot g_j : i \leq m, j \leq n\}$ is a basis for G as a VS over E. Suppose $g \in G$. Express it as $g = \sum_{j=1}^{n} \lambda_i \cdot g_i$, where the λs are in F. Each λ_j can be expressed as $\sum_{i=1}^{m} \mu_{ij} \cdot f_i$ where the μs are in E (co's F is a VS over E). So g was

$$\sum_{i=1}^{m}(\sum_{j=1}^{m} \mu_{ij} \cdot f_i g_j)$$

So the basis spans the whole of G. Just need to check that it is free.

Suppose not...

If $f \in E[X]$ is irreducible then (f) is a maximal ideal in $E[X]$.

Beco's of the division algorithm every equivalence class in $E[X]/(f)$ has a unique representative, namely the remainder on division by f! So we can think of the equiv classes concretely. (It's a pretty faithful implementation)

If f was irreducible of degree 1 you don't add anything, so not much to be gained by thinking of degree 1 polys as irreducible.

Let us now consider the case where F is $E(\alpha)$ for some $\alpha \in F$. ($[F : E]$ not assumed to be finite). Define $\phi : E[X] \to F$ by "evaluate at α". This is obviously a ring homomorphism so let's think about ϕ"$E[X]$. It's a subring of a field and all subrings of fields are integral domains – obviously! Now if the quotient over the ideal is an integral domain, then the ideal was prime (forgotten what prime is!)

[At some point in this discussion α has somehow become the element of the big field corresponding to the identity polynomial over the little field. It turns out that it's a root of the poly in the big field. In some sense one can't tell which one, as the big field has lots of automorphisms fixing the little field. Lots of embeddings from the little field into the big one, presumably as many as there are roots...?

Everything in the big field is a poly over α ...is that right...?]

The powers of α form a basis for the big field as a vector space over the little field...?

Anyway, the upshot is that $\ker(\phi)$ is (f) where either (i) f is 0 or (ii) is irreducible.

In the first case we say α is **algebraic** over E, and in case (ii) it is **transcendental**.

(the old use of "transcendental' is 'transcendental over $\mathbb{Q}$')

(note here: the algebraic closure of a finite field must be infinite. A finite field must be a a finite degree extension of $\mathbb{Z}/p\mathbb{Z}$ and so everything in it must be the root of a poly. So multiply all the polys for the elements together and add 1. The root of this poly isn't in the field.

Something here i'm not getting: Surely there are only finitely many distinct polys over a finite field?. If F is of char p we must have $x^{p-1} = 1$ for all $x \in F$ must we not? No, not quite. Use Lagrange's thm. Multiplicative group of nonzero elements is abelian. Let N be the LCM of the orders of all group elements. Want $n = p-1$. $a \neq 0$ then order$(a)|N$ so $a^N = 1$. $x^N - 1$ has a as a root, so $x - a$ divides this poly. But the ring of polys has unique factorisation, and every nonzero a' is a root so N must be the number of elements. So the multiplicative group is cyclic. So $a^N = 1$ not $a^{p-1} = 1$.

Euclid's proof shows there are infinitely many irreducibles. The point is that once you've added 1 you don't have a unit.)

7.3.1 The Transcendental Case

E a subfield of F, $\alpha \in F$, $F = E(\alpha)$, $\ker(\phi) = \{h \in E[X] : h(\alpha) = 0\} = (0)$.

Ker$(\phi) = \{0\}$, so ϕ is injective, so ϕ"$E[X]$ is iso to $E[X]$ which cannot be a field (rings of polys never are) so ϕ"$E[X]$ cannot be the whole of F. (where does that lead?) Anyway, we can always form the field of fractions over an integral domain and we always get a field. The field of fractions is of course just the field of rational functions with coefficients in E. It seems that we write it $E(X)$.

Now define $\psi : E[X] \to F$ by "evaluate at α". This is OK: α is not algebraic so the denominator is never 0 so no worries about division by zero. Routine to check it's a homomorphism. It's also injective: something like: $g(\alpha)/h(\alpha)$ can only be 0 if g is identically zero co's α cannot be a root of g, being transcendental). α is certainly in the range of ψ co's it's ψ of the identity polynomial (Is it?). So ψ"$E(X)$ is a subfield of F containing α and must therefore be F.

So all fields obtained from E by adding a transcendental are iso! Striking but obvious. Or do i mean obvious but striking?

If α is transcendental, then the powers of α form a basis for F as a vector space over E.

THEOREM 5 *Let F be a finite-degree extension of E. Then*

(i) Every $\alpha \in F$ is algebraic over E;
(ii) $(\forall \alpha \in F)$(deg(minimal polynomial for α) divides $[F : E]$).

Proof of (ii): $E \subseteq F(\alpha) \subseteq F$ So $E(\alpha)$ is a finite dimensionals vector space over E, so α is algebraic over E. Now by multiplicity of degrees of field extensions $[F : E] = [F : E(\alpha)] \cdot deg$(minimal poly for α).

Additional remark: If a, b, c and d are in E, the map $x \mapsto (ax + b)/(cx + d)$ is an automorphism of $E(x)$. Harder is the observation that all automorphisms are of this form. Composition of autos is just matrix multiplication so we get a representation of GL_2(E) acting on $E(X)$.

So α is algebraic over E iff $E(\alpha)$ is a finite degree extension of E. The **minimal polynomial for** f is the unique monic irreducible poly $f \in E[X]$ s.t. $\{h \in E[X] : h(\alpha) = 0\} = (f)$. It's written m_α or m_α^E.

So

$$E[X]/(m_\alpha^E) \simeq E(\alpha); \quad [E(\alpha) : E] = deg(m_\alpha^E)$$

THEOREM 6 *Let E be a subfield of F, α, β both algebraic over E. $E(\alpha, \beta)$ is a finite degree extension of E and $[E(\alpha, \beta) : E] \leq [E(\alpha) : E] \cdot [E(\beta) : E]$*

Proof:

α is algebraic over E sp $E \subseteq E(\alpha)$ so β is algebraic over $E(\alpha)$. Let f be m_α^E and g be $m_\beta^{E(\alpha)}$. Now we exploit the fact that $[E(\alpha) : E] = deg(m_\alpha^E)$ to infer

$$[E(\beta) : E] = deg(f) \text{ and } [E(\alpha, \beta)) : E(\alpha)] = deg(g).$$

Now f and g are both in $E(\alpha)[X]$. $f(\beta) = 0$. g is a generator of the ideal of polys h such that $h(\beta) = 0$. So g divides f in $E(\alpha)[X]$ so $\deg(g) \leq \deg(f)$.

Now $E \subseteq E(\alpha) \subseteq E(\alpha, \beta)$ So $[E(\alpha, \beta) : E] = [E(\alpha, \beta) : E(\alpha)] \cdot [E(\alpha) : E]$ which is $[E(\alpha) : E] \cdot \deg(g) \leq [E(\alpha) : E] \cdot \deg(f) = [E(\alpha) : E] \cdot [E(\beta) : E]$

And there is an easy generalisation to:

If $E \subseteq F$ and $\alpha_1 \ldots \alpha_n$ are in F and all algebraic over E then $E(\alpha_1 \ldots \alpha_n)$ is a finite degree extension of E and $[E(\alpha_1 \ldots \alpha_n) : E] \leq \prod_{i \leq n} [E(\alpha_i) : E]$.

DEFINITION 4
If $E \subseteq F$ the algebraic closure of E in F is $G = \{\alpha \in F : \alpha$ is algebraic over $E\}$.

THEOREM 7

1. *It's a field. Not blindingly obvious, after all. Why should the sum of two surds be a root of a poly with rational coefficients?*
2. *Algebraic closure is idempotent.*

Proof:

1. Obviously $E \subseteq F$. The hard part is the closure under the field operations. $\alpha + \beta$ is in $E(\alpha, \beta)$. But we have just shown that this is a finite degree extension of E so $E(\alpha + \beta)$ (which is a subset of $E(\alpha, \beta)$) is a finite degree extension of E so $E(\alpha + \beta) \subseteq G$. $\alpha \cdot \beta$. This should impress on the reader the advantage of using the vector space analysis! Of course the minimal poly can be computed directly.
2. Let $\gamma \in F$ be algebraic over G with $\sum_{i=0}^{n} \gamma^i \cdot \alpha_i = 0$, not all α_i zero. The α_i are all algebraic over E so $E(\alpha_1 \ldots \alpha_n)$ is a finite degree extension of E by the easy generalisation. Now γ is algebraic over $E(\alpha_1 \ldots \alpha_n)$ so $E(\alpha_1 \ldots \alpha_n, \gamma)$ is a finite degree extension of E and is a subset of G as desired.

■

THEOREM 8
Let $E \subseteq F$, $\alpha \in F$ algebraic over E, and $f \in E[X]$ a monic irreducible such that $f(\alpha) = 0$. Then $f = m_\alpha^E$.

Proof:

$f(\alpha) = 0$, so m_α generates $\{h \in E[X] : h(\alpha) = 0\}$ so m_α divides f in $E[X]$. Dividing irreducibles in Integral domains is a trivial exercise: one factor is always a unit. (Units are the constant polys in this case). And it isn't a unit! ■

How do we spot irreducibles? Let's start with an easy case: how do we recognise irreducible polys in $\mathbb{Q}[X]$?

Any $f \in \mathbb{Q}[X]$ has an associate in $\mathbb{Q}[x]$ which is a primitive element of $\mathbb{Z}[X]$: multiply by the LCM of the denominators. A primitive (check this: i think this is a poly whose coefficients have no common factor) in $\mathbb{Z}[X]$ is irreducible in $\mathbb{Q}[X]$ iff it is irreducible in $\mathbb{Z}[X]$. So it suffices to have a test for irreducibility in $\mathbb{Z}[X]$.

Eisenstein's criterion. Suppose $f \in \mathbb{Z}[X]$ is monic of degree $n > 0$, so that $f = x^n + \sum_{i=1}^{n-1} c_i \cdot x^i$. Let p be prime and suppose that for all i, $p|c_i$ but $p^2 \not| c_0$. Then f is irreducible in $\mathbb{Z}[X]$.

(from memory). Suppose $f = g \cdot h$. The case where one of them has degree 0 is not interesting, since f is monic. So suppose $deg(g)$ and $deg(h)$ are both nonzero.

There is an obvious homomorphism from $\mathbb{Z}$ to $\mathbb{Z}/p\mathbb{Z}$, and we can lift this to a homomorphism $\psi : \mathbb{Z}[X] \to (\mathbb{Z}/p\mathbb{Z})[X]$ coordinatewise. (standard move). Now what does ψ do to f and g and h? Well, it sends f to the poly x^n, so it must send g and h to things that divide x^n. The leading coefficients of $\psi(g)$ and $\psi(h)$ must both be 1 or both -1, so we can take them both to be 1. The ring $(\mathbb{Z}/p\mathbb{Z})[X]$ is a ring of polys over a field so has unique factorisation, so $\psi(g)$ and $\psi(h)$ must be things like x^k. (might need to go over this again). Now think about the constant terms of g and h. They get killed by the mod p homomorphism, and so must be divisible by p. But then the constant term of f must be divisible by p^2.

HIATUS?

The case of $x^2 + 2x + 4$ shows that Eisenstein's criterion is not necessary.

HIATUS

THEOREM 9 *If E is a field, $f \in E[X]$, and $f \neq 0$ then f has at most $deg(f)$ roots in E.*

Proof: Well, if E is merely a ring then there might be more than $\deg(f)$ roots. $x^2 - 1 = 0$ has *three* roots in $\mathbb{Z}/8\mathbb{Z}$ – 1, 3 and 7. If $f(a) = 0$ then $(x - a)|f$ in $E[X]$ which is a UFD. ■

Consider in $C[X]$ the poly $x^n - 1$, with $n \in \mathbb{N}$, $n > 1$. The roots of this poly in $\mathbb{C}$ are $\{e^{(2\pi i)/m} : 1 \leq m \leq n\}$, beco's if a is a root we must have $a = r \cdot e^{i\theta}$ for some r and θ. So $a^n = r^n \cdot e^{ni\theta}$. Now $|e^{ni\theta}| = 1$ and $|z \cdot w| = |z| \cdot |w|$ so $r = 1$ and $n\theta$ is a multiple of 2π.

Complex conjugation is an automorphism of $\mathbb{C}$, so polys with real coeffs must have conjugate pairs of roots! (Fix the poly you fix the set of roots!!). This (very cute) argument underlies Galois theory.

A cyclotomic field is a field of the form $\mathbb{Q}(e^{(2\pi i)/n})$ for some $n \in \mathbb{N}$. ("circle cutting fields"). Obviously something to do with Fermat's conjecture.

Q: what is the minimal polynomial of $e^{(2\pi i)/n}$ over $\mathbb{Q}$? (why isn't it just $(x^n-1)/(x-1)$? or $(x^n-1)/(x^2-1)$? Beco's we care which of the factors of the quotient poly is the one of which *that particular root* is a solution). But it does at least *divide* x^n-1. Let us start by knocking off the easy case where the exponent is an odd prime: x^p-1 is irreducible.

Consider the operation sending a poly g to $[(x+1)/x]g$ - the result of substituting $x+1$ for x in g. This is an automorphism of $C[X]$ that fixes all constant polys. beco's it's an automorphism it preseves irreducibility. The new poly is $((x+1)^p-1)/(x+1-1)$ so all its coefficients are divisible by p but the constant term isn't. Now that's what i call cute.

7.3.2 Compass and straightedge constructions

Think of compass and straightedge constructions as a way of generating new points in the plane from old. We are allowed the following operations:

1. Given distinct points A and B can draw the (infinite) line thru' them;
2. Given two points A and B can put compass point at A and the pencil at B and draw a circle thru' B;
3. Add intersection of two lines to your set of points;
4. "Copy" lengths obtained in step (2).

Given a set $Y \subseteq \mathbb{R}^2$ define

$$Q(Y) =: \{a \in \mathbb{R} : a \text{ is the ordinate or abcissa of a point in } Y\}.$$

When we perform (3) above all we are doing is solving a quadratic equation with coefficients in the field so far. So if we start with a set of points with rational coefficients we will get a set Y of points s.t. $[Q(Q(Y)) : Q]$ is a power of two. (overloading of Q!!)

Now if $m > 2$ is divisible by an odd prime p not of the form $2^n + 1$ then it is not possible to start with the two points $(0,0)$ and $(0,1)$ amd construct a regular m-gon. If you could, you could construct a regular p-gon but the degree of the cyclotomic polynomial is $p-1$ which by hypothesis is not a power of 2.

7.3.3 Galois Groups

The **Galois Group** $\Gamma(F/E)$ of a field F over E is the group of those automorphisms of F that fix everything in E. E.g., $\Gamma(\mathbb{C}/\mathbb{R}) = \{\text{conjugation, identity}\}$. This is worth proving. $\mathbb{C}$ is $\mathbb{R}(i)$ so if τ is an automorphism it must fix the set

of roots of the characteristic poly $x^2 + 1$. (since it fixes the coefficients). But once we know what it does to i we know what it does to everything, since we've determined τ on a basis for $\mathbb{C}$ as a VS over $\mathbb{R}$.

THEOREM 10 *If $[F : E]$ is finite, $|\Gamma(F/E)| \leq [F : E]$*

Equality doesn't always hold. (don't know how to do cube roots in LaTeX!) Think about automorphisms of $\mathbb{Q}(2^{1/3})$ that fix everything in $\mathbb{Q}$. Clearly there is only the identity. This isn't sensible. We should have thought about automorphisms of $\mathbb{R}(2^{1/3})$ that fix everything in $\mathbb{R}$.

(Notice that this is nontrivial: the obvious bound that one gets from the *aperçu* that every automorphism arises from a permutation of the roots is $\deg(f)$!. So we should expect to have to do some work.)

A "character of G in F", for G a group and F a field, is a group homomorphism from G to the multiplicative group of nonzero elements of F. They form a group under pointwise multiplication.

EG, a character of integers-mod-3-with-addition in $\mathbb{C}$ is either $\lambda x.1$ or
$0 \mapsto 1; 1 \mapsto e^{2\pi i/3}; 2 \mapsto e^{4\pi i/3}$ or
$0 \mapsto 1; 1 \mapsto e^{4\pi i/3}; 2 \mapsto e^{2\pi i/3}$.
the last two arise from "send a generator of the group to an element of order 3"

Let us use the Greek letter χ to range over characters.

Artin: there is a natural notion of F-linear combination of characters, elements of the VS of functions $G \to F$ which is naturally a VS over F.

THEOREM 11 *(Artin)*
The set of characters is a linearly independent set

(but is not a basis unless G is small-and-nice and F is big (eg if G is finite abelian and $F = C$. Watch this space)
Proof:

Suppose $\lambda_1 \dots \lambda_n \in F$ and for all g in G, $\sum_{i=1}^{n} \lambda_i \chi_i(g) = 0$, then the λ_i are all 0. (I think for the λ_i to be all zero we have to assume that this equation holds for all $g \in G$: ie, for all tuples of λ, if for all $g \dots$)

Our proof is by *reductio.* Let $\chi_1 \dots \chi_n$ be all distinct. Suppose there is a n-tuple of λs and shrink it to an $n-1$-tuple of λs. WLOG the λs are all nonzero.

$\chi_1 \neq \chi_n$ so $\exists g \in G\ \chi_1(g) \neq \chi_n(g)$. Now for any $h \in G$, $gh \in G$ so

$$\sum_{i=1}^{n} \lambda_i \cdot \chi_i(gh) = 0$$

so

$$A : \sum_{i=1}^{n} (\lambda_i \cdot \chi_i(g)) \cdot \chi_i(h) = 0$$

which is a new linear dependence. Also $\sum_{i=1}^{n} \lambda_i \cdot \chi_i(h) = 0$ so multiply this last equation by $\chi_n(h)$ to get

$$A : \sum_{i=1}^{n} (\lambda_i \cdot \chi_n(g)) \cdot \chi_i(h) = 0$$

Now subtract B from A to get

$$A : \sum_{i=1}^{n-1} (\lambda_i (\chi_i(g) - \chi_n(g))) \cdot \chi_i(h) = 0$$

We want at least some of the coefficients to be nonzero. But we took care to use a g s.t. $\chi_1(g) \neq \chi_n(g)$. ■

An automorphism of F gives a character of F in F! So the family of automorphisms of F is linearly independent over F.

We now resume the proof that if $[F : E]$ is finite, $|\Gamma(F/E)| \leq [F : E]$. Let n be $[F : E]$ and fix $\langle f_1 \dots f_n \rangle$ a basis for F as a VS/E.

Suppose for a contradiction that $\sigma_1 \dots \sigma_{n+1}$ are distinct elements of $\Gamma(F/E)$. Recall that F^n is a VS of dimension n over F. We define elements $v_1 \dots v_{n+1}$ of F^n where v_i is the column vector $\sigma_i(f_1) \dots \sigma_i(f_n)$. [*HOLE do column vectors in LaTeX*]

$v_1 \dots v_{n+1}$ must be linearly dependent, so we can find $\lambda_1 \dots \lambda_{n+1}$, not all zero, such that $\sum_{i=1}^{n+1} \lambda_i \cdot v_i = 0$). Equivalently

$\sum_{i=1}^{n+1} \lambda_i \cdot \sigma_i(f_j) = 0$ for all j with $1 \leq j \leq n$. So this linear combination of the σs vanishes at each basis element! So we won't be surprised to learn that it vanishes everywhere. Let's just check this.

Suppose $f = \sum_{j=1}^{n} \mu_i \cdot f_j$ with $\mu_j \in E$

$$\sum_{i=1}^{n+1} \lambda_i \cdot \sigma_i(f) = \sum_{i=1}^{n+1} \lambda_i \cdot \sigma_i(\sum_{j=1}^{n} \mu_i \cdot f_j)$$

co's σ is an automorphism

$$= \sum_{i=1}^{n+1} \lambda_i \cdot (\sum_{j=1}^{n} \sigma_i(\mu_j) \cdot \sigma_i(f_j))$$

$$= \sum_{i=1}^{n+1} \lambda_i \cdot (\sum_{j=1}^{n} \mu_j \cdot \sigma_i(f_j))$$

(co's σ fixes everything in E)

$$= \sum_{j=1}^{n} \mu_j \cdot (\sum_{i=1}^{n+1} \sigma_i(\lambda_i) \cdot \sigma_i(f_j))$$

$$= \sum_{j=1}^{n} \mu_j(0) = 0$$

So we've produced a nontrvial linear dependence between the sigmas. ■

Context: $\Gamma(F/E)$ finite

DEFINITION 5 *Let $f \in E[X]$. Say f* **splits over** *E iff it is a product $c \cdot \Pi_{i=1}^{n}(x - a_i)$ with c and all the a_i in E.*

(So "f splits over E" doesn't make sense unless the coeffs of f are in E.)

DEFINITION 6 *Let E be a subfield of F, $f \in E[X]$. Then F* **is a splitting field for** *f* **over** *E iff*

1. *f splits over F*
2. *$F = E(a_1 \dots a_n)$, the roots of f.*

The splitting field for f over E is the most economical field that splits f in the sense that $E \subseteq G \subseteq F$ and f splits over G and F a splitting field for f over E then $G = F$.

Splitting fields always exist and they are unique.

Existence:

THEOREM 12 *If E is a field and $f \in E[X]$, $deg(f) = n$ then $\exists F$ a splitting field for f over E, and $[F : E] \leq n!$ (and $n!$ is the best we can do)*

[stuff missing here]

7.3.4 Galois extensions

DEFINITION 7 *If F is a finite degree extension fo E then F is a* **Galois** *extension of E iff $(\forall \sigma \in \Gamma(F/E))(\sigma(a) = a)$*

DEFINITION 8 *For $X \subseteq Aut(F)$ let $Fix(X)$ be $\{a \in F : (\forall \sigma \in X)(\sigma(a) = a)\}$*

THEOREM 13 *$Fix(X)$ is always a subfield.*

(No conditions on X!)

Examples. $\mathbb{C}$ is a Galois extension of $\mathbb{R}$; $\mathbb{Q}(\sqrt{2})$ is a Galois extension of $\mathbb{Q}$; $\mathbb{Q}$ of cube-root-3 is not a Galois extension of $\mathbb{Q}$: the Galois group is trivial.

THEOREM 14 *Let F be a field, H any finite subgroup of $Aut(F)$, and $E = Fix(H)$. Then*

1. *F is a finite degree extension of E;*
2. *F is a Galois extension of E and $\Gamma(F/E) = H$.*

Proof:

Start by noting that H is a subgroup of $\Gamma(F/E)$.

Key claim. F is a finite degree extension of E and $[F:E] \leq |H|$. Suppose not, and let n be $|H|$ and find $n+1$ elements of F which are linearly independent over E, call them $f_1 \dots f_{n+1}$. Enumerate H as $\sigma_1 \dots \sigma_n$ where σ_1 is the identity.

If $n = 1$ then $E = F$, $[F:E] = 1 = |H|$ so we're OK. For each $1 \leq j \leq n+1$ define a vector $w_j \in F^n$ as $\sigma_1(f_j) \dots \sigma_n(f_j)$. F^n is a VS/F of dim n so there is a linear relation between the w_j. Now let k be the smallest size of a linear dependence. The f_i are all nonzero, so $k > 1$. Relabelling if necessary we may assume that $\{w_i \dots w_k\}$ are linearly dependent. Next choose $\lambda_1 \dots \lambda_k$ so that $\sum_{j=1}^{k} \lambda_j w_j = 0$. The λs are all nonzero, by minimality of k. Multiply everything by $(\lambda_1)^{-1}$ and abbreviate $(\lambda_1)^{-1} \cdot \lambda_j$ to μ_j. This gives $\sum_{j=1}^{k} \mu_j w_j = 0$ and $\mu_1 = 0$. (Surely it should be $\mu_1 = 1$..?) Therefore, applying τ, for any $\tau \in H$,

$$(*)_\tau \qquad \sum_{j=1}^{k} \mu_j(\tau(f_j)) = 0.$$

In particular, the identity is in H so $\sum_{j=1}^{k} \mu_j \cdot f_j = 0$. (We haven't yet used the fact that H is a group). Now the f_js are independent over E, so not all the μ_j are in E (Note that $\mu_1 \in E$!) Relabelling, we may assume that $\mu_k \notin E$. Since $E = FIX(H)$ we can find $\sigma \in H$ s.t. $\sigma(\mu_k) \neq \mu_k$. Now apply *this* σ to $(*)_\tau$ to get

$\sum_{j=1}^{k} \sigma(\mu_j)(\sigma \circ \tau(f_j)) = 0$, and this of course holds for every $\tau \in H$.

The field of rational functions is differentially closed! Think about the polynomial notation for $\mathbb{N}$. It's a map $\mathbb{N} \hookrightarrow \mathbb{Z}/10\mathbb{Z}[X]$

7.4 Number fields

A Number field is an extension of $\mathbb{Q}$ of finite degree, and a subfield of the complexes (inevitably). So an **algebraic integer** is a root of a monic polynomial with integer coefficients. An algebraic integer that happens to be rational is actually going to be in $\mathbb{Z}$. (This shouldn't be too hard to prove) hence **rational integer**.

We characterise them, and prove that they are a ring. Suppose a is complex, $Z[a]$ is the least subring of $\mathbb{C}$ containing a. Consider the additive group structure of $Z[a]$. a is an algebraic integer iff that is a finitely generated group.

L $\rightarrow$ R

Suppose a satisfies a monic polynomial of degree n. Then a^n can be expressed in terms of smaller powers of n

R $\rightarrow$ L

Fix some generators. Express each generator as a poly in a with integer coefficients. The pick $n >$ degree of any of those polys. Consider a^n in terms of the generators. This is a monic polynomial of degree n. (Need to polish this up – a bit obscure...)

Now suppose a, b are algebraic integers. $\mathbb{Z}[a, b]$ is finitely generated (becos $\mathbb{Z}[a]$ and $\mathbb{Z}[b]$ both are). $\mathbb{Z}[a + b] \subseteq \mathbb{Z}[a, b]$.

So algebraic integers are closed under $+$.

Let $F \subseteq \mathbb{C}$ be an extension of $\mathbb{Q}$ of finite degree. So F is a number field, and is $\mathbb{Q}(\alpha)$ for some α. Think of the ring of integers of F, which is the set of algebraic integers that are in F. It turns out that this is $\mathbb{Z}(\alpha)$

It's a Unique factorisation domain iff it's a Principal ideal domain. (hard) PIDs have a notion of HCF. HCF(a, d) = a generator of (a, b). Thus the 'ideal' generated by a and b is a ficticious HCF. Something to do with $f \in (g)$ iff g divides f. Obvious!

An ideal I in a ring R is prime iff R/I is an integral domain. James sez: the two clauses of the dfn of prime ideal correspond to the two clauses in the dfn of integral domain. $x \cdot y \in I \to .x \in I \vee y \in I$ takes care of no divisors of zero. The properness condtion ensures that $0 \neq 1$.

Prime ideals in boolean rings are all maximal beco's the only BA that is an integral domain is the two-element BA.

7.5 Notes on JWSC's Part III course on local fields

James sez: Given a field with a metric, you can complete it wrt the metric or you can close it algebraicly. By the time you've done both, does it matter which way round you did it? Not with $\mathbb{Q}$. It does with the p-adic rationals.

All started by Hensel.

Concept of *absolute value* of things in (i think) a **field**. written $|x|$. It must satisfy the following conditions:

Definition 9

(i) $|x|$ *is always a nonegative real, and* $|x| = 0 \to x = 0$*;*

(ii) $|(x \cdot y)| = |x| \cdot |y|$*;*

(iii) $|a + b| \leq |a| + |b|$.

This is all we usually use. (The condition (iii) says that $|a - b|$ is a metric.)

Let K_0 be a field and $K_0(T)$ be the new field obtained by adding one chap T; call it K for short. (Think of ring of polynomials in T with coefficients in K_0) Fix a constant $c > 1$. Then – if f is a polynomial of degree n – set $|f| =: c^n$.

Check that this satisfies (i) - (iii). Indeed it even satisfies

(iii)*: $|f + g| \leq \max(|f|, |g|)$.

(iii)* implies $|a + b| = |a|$ if $|b| \leq |a|$.

Extend this to rational functions by declaring

$$|f/g| =: |f|/|g|$$

Check that this is uniquely defined.

7.5.1 p-adics

Look at the field $\mathbb{Q}$ and fix a prime p. Every $r \in \mathbb{Q}$ is $p^\rho \cdot \frac{u}{v}$ with p not dividing u or v, u and v coprime. ρ is unique. Then the p-adic value of r – written "$|r|_p$" – is $p^{-\rho}$. (perhaps we mean that the value is ρ)

Check that is satisfies clauses (i) – (iii) of definition 9: (i) – it's not only real-valued, it's integer valued (No it isn't: what did i mean???) (ii) OK. (iii) is more like hard work

Suppose $s = p^\sigma \cdot \frac{x}{y}$ with $x, y \in \mathbb{Z}$, $p \not| xy$, and similarly $r = p^\rho \cdot \frac{u}{v}$ with $u, v \in \mathbb{Z}$, $p \not| uv$. Without loss of generality $\sigma > \rho$ with $\sigma = \rho + \tau$ so

$$r + s = p^\rho(\frac{uy + p^\tau xu}{vy})$$

where $p \not| vy$.

However p might divide the numerator. Suppose it does. Then the numerator is $p^{\rho+\lambda} \cdot \frac{l}{m}$ with $\lambda \geq 0$ and $p \not| lm$. This gives

$$|r + s| = p^{-\rho-\lambda} \leq p^{-\rho} = |r| = \max\{|r|, |s|\}$$

This is the **p-adic valuation**.

Check that $|1| = 1$. [*HOLE but it isn't! $1 = p^0$ ...*]

7.5.2 Completing the rationals

Now think of limits. All we use in $\mathbb{R}$ is (i) – (iii) which we have here so we can think about completing the rationals with respect to this metric.

A brief digression to accomodate a Conwayism. Set $p =: 5$ and consider the sequence

$$a_1 =: 3; \quad a_{n+1} = 3 + 10 \cdot a_n$$

namely 3, 33, 333, 3333, ...

We have $3 \cdot a_n = 10^{n+1} - 1$, so $|3 \cdot a_n + 1| = 5^{-n-1}$ which gets very small, so $3 \cdot a_n$ gets arbitrarily close to -1, and a_n tends to $-1/3$. [*HOLE Explain this!*]

Now the rationals with the p-adic topology aren't an ordered set so we have to use Cauchy sequences rather than Dedekind cuts to complete it. (Clarify this)

A *fundamental (Cauchy)* sequence $\langle a_n : n \in \mathbb{N} \rangle$ satisfies

$$(\forall \epsilon > 0)(\exists n_0)(\forall m, n > n_0)(|m - n| < \epsilon).$$

A space is **complete** if every sequence has a limit. [*HOLE better explain what this is, for the sake of ... completeness ha ha!*]

Look again at the case $p =: 5$. Construct a sequence $\langle a_n : n \in \mathbb{N} \rangle$ such that

$$(a_n)^2 + 1 = 0 \ (\mathrm{mod} 5^n)$$

Try $a_0 =: 2$. OK 'co's $2^2 + 1 = 5$. Thereafter set $a_{n+1} =: a_n + 5^n \cdot t$ where t is to be determined later.

Then $(a_{n+1})^2 = {a_n}^2 + 2 \cdot 5^n \cdot a_n + 5^{2n} \cdot t^2$.

Therefore $|a_{n+1} - a_n| = 5^{-n}$ and $\langle a_n : n \in \mathbb{N} \rangle$ is a fundamental sequence. Suppose it had a limit, b. Then, as $n \to \infty$, $a_n \to b$ and – since products and sums commute with limits – ${a_n}^2 + 1 \to b^2 + 1$ but $a^2 + 1 = 0$ [*HOLE What did i mean?*] so $b^2 + 1 = 0$. But this is not possible.

7.5.3 Valuations

A field K, with a real valued function $||$ on K such that

(i) $|x|$ is nonnegative and is 0 iff x is too.

(ii) $||$ is multiplicative: $|a \cdot b| = |a| \cdot |b|$.

(iii) $\exists c$ depending only on K such that $|a| \leq 1 \to |1 + a| \leq c$

Corollaries:

(iv) $|1_K| = 1_{\mathbb{R}}$. (Think of $|1_K|^2$.)

(v) $(a^n = 1_k) \to |a| = 1_{\mathbb{R}}$.

The trivial valuation sends all nonzero chaps to $1_{\mathbb{R}}$. By (v) the only valuation on a finite field is the trivial valuation.

(vi) If α is a real > 0 then $\lambda x_K.|x|^\alpha$ is a valuation too.

(vii) Every valuation is equivalent to one satisfying the triangle inequality. [*HOLE What is "equivalent"?*] We can suppose $c = 2$. Without loss of generality [*HOLE why?*] and we will show that this implies the triangle inequality.

$|a+b| \leq max(|a|, |b|)$ and without loss of generality $|b| \leq |a|$ so $b = a \cdot c$ with $c \leq 1$. [*HOLE doesn't this $\leq$ mean K must be an ordered field, and nobody said nuffin' about that!?*]. Then $|a + b| = |a| \cdot |1 + c| \leq c \cdot |a|$.

Then, by induction,

$$|a_1 + a_2 + \ldots a_{2^n}| \leq 2^n \cdot max_{i \leq 2^n} |a_i|$$

and

$$|a_1 + a_2 + \ldots a_N| \leq 2N \cdot max_{i \leq N} |a_i|.$$

Now $N = 1 + 1 + 1 + 1 \ldots$, N times.

Consider $|a + b|^n = |(a + b)^n|$.

Expand using binomial theorem. This is

$\leq 2(n + 1) maxterm$ which is

$\leq 2(n + 1) max_j 2\binom{n}{j} |\alpha|^j \cdot (|a| + |b|)$

replace max by sum

$\leq 4(n + 1)(|a| + |b|)^n$

All in the reals, so take nth root, getting

$|a+b| \leq (4 \cdot (n+1))^{1/n} \cdot |a| \cdot |b|$
Now let $n \to \infty$ getting
$|a+b| \leq |a| + |b|$.
So WLOG assume triangle inequality.

Definition 10 *Say a valuation is* **non-archimedean** *if we can take $c = 1$ in (iii) above. Otherwise it is archimedean.*

K is archimedean iff everything in the ring generated by $\{0_K, 1_K\}$ has value ≤ 1.

One direction is easy. For the other, remember that this is a property of the equivalence class.

Corollary 3 *If K is not of characteristic 0 all valuations of K are nonarchimedean.*

Corollary 4 *Suppose $k \subset K$. Any valuation on K restricts to a valuation on k, and it will be archimedean on the big field iff it's archimedean on the little field.*

We are going to classify all valuations on $\mathbb{Q}$. It will turn out that they are either the trivial, the usual, or one of the p-adic valuations. p-adic valuations for distinct p are inequivalent.

Consider a valuation on $\mathbb{Q}$. Without loss of generality it satisfies the triangle inequality. Let $a > 1$ be in Z, and $b > 0$. Express b as a sum of multiples of powers of a thus:

$$b = b_m \cdot a^m + b_{m-1} \cdot a^{m-1} + b_{m-2} \cdot a^{m-2} \dots$$

where $0 \leq b_j \leq (a-1)$, $b_m \neq 0$ and $m \leq \frac{log\ b}{log\ a}$.

Now apply the triangle inequality. $|b| \leq$ sum of values on the right hand side. Th

$|b| \leq A \cdot max(1, |a|^m) \dots$ where $A = max\{|0|, |1|, |2|, \dots |(a-1)\}$ (independent of b)

7.6 Chris Brookes on Part II Galois theory; michaelmas 2017

Lagrange's theorem was something to do with Lagrange trying to see why the algorithms for solving cubics and quartics worked. 1799 Ruffini claimed there are quintics not solved by radicals, but the proof had gaps. Abel proves it properly in 1824; Galois *explains* it. He considers permutations of the set of roots, and understood the importance of normal subgroups.

[It's just struck me: quadratics cubic and quartics soluble by radicals. This means not only that for any poly of degree ≤ 4 there is an expression over the language $\sqrt{}$, $+$, $\times$ which denotes a root; it means that it can be done uniformly

and simultaneously for all polys of that degree: there is a single formula in the coefficients with the property that if you stick in values for the coefficients you get out the values of the roots. The unsolvability of the quintic means not only that there is no *uniform* solution; it means that there a quintic with no solution in radicals.]

If $K \leq L$ is a field extension (as they so quaintly put it) then L can be thought of as a vector space over K. Chris didn't say how, so presumably it's obvious, but i need to spell this out for my own benefit.

Write '$|M : K|$' ("the degree of M over K") for the dimension of M as a vector space over K.

Then

THEOREM 15 *The Tower Theorem*

If $K \leq L \leq M$ and all the degrees are finite then

$$|M : K| = |M : L| \cdot |L : K|$$

Proof:

You pick a basis for L as a VS over K and a basis for M as a VS over L and this gives you a basis for M over K. Then you have to prove that this last basis is genuinely an independent set. ■

For example $\mathbb{Q} \leq \mathbb{Q}[\sqrt{2}] \leq \mathbb{Q}[\sqrt{2}, i]$. $\mathbb{Q}[\sqrt{2}, i]$ has degree 4 over $\mathbb{Q}$, with basis $\{1, \sqrt{2}, i\}$. (Degree of M over L is 2^{b-1} where b is the size of the basis of M as a VS over $L \ldots ?$) So clearly if we have a field K with $\mathbb{Q} \leq K \leq \mathbb{Q}[\sqrt{2}, i]$, then the degree of K over $\mathbb{Q}$ must be a factor of 4. **How many such fields are there??**

7.6.1 Some observations on the quintic from Jeroen Schillewaert NZMA Dec 2023

We consider the set of quadruples $[\mathbb{C}]^4$ of complex numbers. His culture has a funny notation for it, but there you go. *Un*ordered quadruples, members are pairwise distinct. Consider the map

$$f : \{x_1, x_2, x_3, x_4\} \mapsto \{x_1x_2 + x_3x_4,\ x_1x_3 + x_2x_4,\ x_1x_4 + x_2x_3\}$$

Useful facts.

- If the x_i are all distinct then so are $x_1x_2 + x_3x_4$, $x_1x_3 + x_2x_4$, and $x_1x_4 + x_2x_3$.
- $f : [\mathbb{C}]^4 \twoheadrightarrow [\mathbb{C}]^3$. This is not obvious but presumably can be checked easily enuff. It is also unique (up to affine something-or-other) with this property.
- Jeroen says he has proved that that f has no holomorphic right-inverse. It is alleged that it doesn't even have a *continuous* right-inverse.

- This function was known to the *18-year old* (!) Ferrari in 1520(!) and it is a crucial gadget in the proof that there is a way of solving quartics in radicals using the fact that cubics can be solved by radicals.

Look up Arnol'd's elementary proof that there is no solution of the quintic in radicals

7.7 A IB GRM lecture from Imre on Sylow's theorem

Let G be a group of cardinality $p^m \cdot a$ with p prime and $(p, a) = 1$. We know from a theorem of Cauchy (in Ia) that G has a subgroup whose size is a power of p (a "p-subgroup"), so let P be a p-subgroup of maximal size. Let us see what we know about such a P. The first thing is to show that $|P| = p^m$. The way to do this is to show that $|G|/|P|$ is not a multiple of p. [say a bit about why that is enuff].

Consider N the normaliser of P. "Normaliser"?! Wossat? It's $\{g \in G : gPg^{-1} = P\}$. Brief digression to check that the operation *normaliser-of* is idempotent. Also $P \trianglelefteq N$ by definition of normaliser. This will matter.

Now comes the first clever move:

$$|G|/|P| = |G|/|N| \cdot |N|/|P|$$

The two factors on the RHS are both meaningful:

(i) $|N|/|P| = |N/P|$ [note overloading of the slash]

(ii) $|G|/|N|$ is the number of conjugate copies of $P \subseteq G$ (by orbit-stabiliser theorem).

We start with (i). We want this quantity *not* to be 0 mod p. Well, we have the quotient map $\pi : N \twoheadrightarrow N/P$. Suppose $|N/P|$ congruent to 0 mod p, then $|P|$ divides $|N/P|$ so there is a subgroup W of order p by Cauchy. Take the preimage π^{-1}"W. This is a subgroup of N (and therefore of G) of order $p \cdot |P|$ contradicting maximality of P.

Now we work on (ii)

Write 'X' for the set $\{gPg^{-1} : g \in G\}$ of conjugate copies of P. (We will be seeing a lot of this set).

The key observation is that P acts on X by conjugation. The fact that $|P|$ is a power of p tells us that the size of a P-orbit $\subseteq X$ is a power of p. There is an orbit of size 1, namely $\{P\}$; the challenge is to show that this is the *only* singleton orbit. To this end, suppose $\{gPg^{-1}\}$ is a singleton orbit. That is to say, $(\forall h \in G)(hgPg^{-1}h^{-1} = P)$. But this says that $gPg^{-1} \subseteq N$. Consider now where this gPg^{-1} gets sent by the quotient map π. π"(gPg^{-1}) must be the trivial subgroup of N/P, co's it's a subgroup of N/P so its order divides $|N/P|$ and $|N/P|$ is not a multiple of p [why?]

OK, so $|P| = p^m$. In other words, the p-subgroups of G of maximal size are all of size p^m. Now we want to show that all these P-candidates (called "Sylow p-subgroups") are mutually conjugate. So, let Q be another P-candidate. Consider *its* action on X (just as we considered the action of P on X earlier). [long story] just like P, Q must fix some element of X.

Finally, how many of the buggers are there? Observe that, by the mutual-conjugacy-of-P-candidates that we have just established, X is not only the set of conjugates of P, but is precisely the set of P-candidates. Let us write 'n_p' (the $n$umber of p-subgroups of maximal size") for '$|X|$'. X is an orbit, so $|X|$ divides $|G|$ which is to say $n_p | p^m \cdot a$. But then $n_p | a$, since $p \not| n_p$. [why?]

Chapter 8

Assorted other topics

8.1 A Talk by Alex Wilkie

Theory of algebraically closed fields is complete and decidable. This means that there is an algorithm that will uniformly solve various standard problems, like for example the dispute between Newton and Gregory (1694) over sphere packing. How many nonoverlapping unit spheres can be tangent to a unit sphere? Obviously at least 12 but at most 12? Gregory said we should be able to do 13, and Newton said no. So consider the formula $\Phi_{3,13}$ which is

$$(\exists x_1 \dots x_{13})(\bigwedge_{1<i<13} |x_i|^2 = 4 \wedge \bigwedge_{1\leq i\leq j\leq 13} |x_i - x_j|^2 \geq 4$$

(i suspect some of the inequalities should be strict or *vice versa*. Not sure)

Of course we can generalise this to higher dimensions. Naturally this is related to packing numbers, Let $N(d) =:$ the largest number of d-dimensional spheres that can be tangent to a given d-dimensional sphere.

8.1.1 Thomas Forster's notes of Alex Wilkie's talk at BLC, Manchester sept 2001. Comments by the auditor enclosed in square brackets.

We are looking at the theory of the reals with $+$, $\cdot$, -, 0, 1 and $\leq$. Call this structure $\mathbb{R}_{alg}$. It's an old result of Tarski that there is a primitive recursive procedure for eliminating quantifiers, whence decidability.

Here is an entertaining example of where this might have been useful. Newton and Gregory argued in 1694 about how many non-overlapping spheres of the same radius can be simultaneously tangent to a given sphere (of the same radius). The answer is obviously at least 12, but the 12 tangent spheres are quite loosely packed. [comment by tf: someone once told me that in the obvious packing any two spheres can be permuted by rolling the spheres around while leaving all the other tangent spheres in their original position. Can't remember

where i found this]. Newton said the number was 12, and Gregory had a hunch that it was 13. Let $\phi_{n,m}$ say that in n dimensions the maximum number of non-overlapping spheres of the same radius that can be simultaneously tangent to a given sphere (of the same radius) is m. Thus Newton and Gregory were arguing about the truth-value of $\phi_{3,12}$ which is

$$(\exists x_1 \dots x_{13})(\bigwedge_{1<i<13} (|x_i|^2 = 4 \wedge \bigwedge_{1\leq i\leq j\leq 13} |x_i - x_j|^2 \geq 4))$$

We could apply Tarski's algorithm to this to get the answer, but as it happened this particular case was settled in 1874 by a chap called Hopper long before Tarski discovered the algorithm. Another proof by Van der Waerden in 1923.

This is of course related to packing problems. Let us define $N(d)$ (the "Newton number" for d) to be the largest number of non-overlapping spheres of the same radius that can be simultaneously tangent to a given sphere (of the same radius) in n dimensions. Thus $N(3) = 12$. $N(d) < 3^d - 1$ beco's all the spheres tangent to the sphere in the middle have to be found inside the sphere of radius 3 concentric with it. There are surprisingly few values known. $N(2) = 6$, obviously, $N(3) = 12$ (so Newton was right!), $N(8) = 240$ and $N(24) = 196,560$. [Alex threw off this last number for a laugh and gave no explanation but i believe it's something to do with the Leech lattice]

$\lambda d.N(d)$ is actually explicitly primitive recursive: given d, just run the algorithm to test whether or not the theory of real-closed fields believes $\Phi_{d,1} \dots \Phi_{d,n}$ until it rebels and says no. That is to say, given d, just use Tarski's algorithm to do the following:

$n =: 0$
`REPEAT` $n =: n + 1$; $\phi(d, n)$
`UNTIL`
`false`

Definable subsets of $\mathbb{R}$ are finite unions of open intervals and singletons, where the singletons and endpoints are real algebraic numbers. Aperçu by Lou van den Dries: this doesn't use the algebraic structure so use it as a definition. We say that a structure with a dense linear order is **o-minimal** iff every definable subset of it is a finite union of open intervals (and singletons??). The 'o' is an Oh not a zero, and it stands for 'order'.

Can do this for hyperbolic geometry if we have exponentials as well. [This sounds a bit cryptic – what did he mean by this exactly?]. So this is a reason to think about the reals with all the stuff they've had so far, plus the exponential function: call this structure $\mathbb{R}_{exp}$. No quantifier elimination for this structure [there's a counterexample due to Lou v.d.D quoted in one of the *Notices* articles.] However every formula is equivalent to a $\sum_1$ formula which by a result of Khovanskii is enuff to imply that $\mathbb{R}_{exp}$ is o-minimal.

A d -cell in M^n is **either** the graph of the restriction to a d-cell in M^{n+1} of a definable cts function, **or** $\{\langle \vec{x}, y\rangle \in M^{n+1} : \vec{x} \in C', h(\vec{x}) < y < g(\vec{x})\}$ where C'

is a $(d-1)$-cell and h and g are definable cts with $h < g$ in some sense obviously required by this definition. 0-cells are singletons; 1-cells are open.

Now we have:

Definable sets of n-tuples of an o-minimal structure are finite unions of cells.

$dim(X)$ = largest d such that X has a subset C that is a d-cell.

There is also an analogue of the Euler number (like: vertices + edges + surfaces = constant) defined as follows [or something like that – what i wrote down actually doesn't make sense]

$$\text{euler}(X) =: \textstyle\sum_0^n (-1)^d n_d$$

(yes, i know, it doesn't make sense)

$n_d(C)$ = number of d-cells in C. Euler number is preserved under cell-decomposition, preserved by definable bijection. Finally if our o-minimal structure has a field structure the converse is true: between two things of the same dimension and the same Euler number there is a definable bijection.

Examples of *o*-minimal structures.

- Additive group of the rationals;
- $\mathbb{R}_{alg}$ and $\mathbb{R}_{exp}$;
- The class of *o*-minimal structures is closed under elementary equivalence;
- Add to $\mathbb{R}_{alg}$ the restriction to an arbitrary open interval U of an analytic function f. This is also *o*-minimal. (One has to restrict it to an open interval – or at least something compact, otherwise by using **sine** and **cos** one could define sets that weren't the unions of finitely many open intervals. [they would be the union of a finitely presented – indeed *recursive* – set of open intervals ...]

Can't amalgamate

8.2 The ABC conjecture

Let $R(x) :=$ product of all the distinct prime factors of x. Someone made the point to me once that we know of no way of computing $R(x)$ without factorising x.

Conjecture:

There is a constant k such that, for all $a+b=c$ (all +ve, coprime), $c \leq R(abc)^k$.

A strong form of the conjecture is that there are only finitely many exceptions for any value of $k > 1$.

We know that $\texttt{log}(c)/\texttt{log}(R(abc))$ can be as much as 1.629912 ($a = 2$, $b = 3^{10} \cdot 109$, $c = 23^5$). There are about 1000 cases with this quantity $\geq 1 \cdot 2$.

Any upper bound on k would imply Fermat's Last Theorem: $z^n \leq R(x^n y^n z^n)^k = R(xyz)^k \leq z^3 k$.

Have i got this right?

For all k > 0 (however small) there is a c (so large that) for all x, y coprime and bigger than c $(R(x.y.(x+y)))^{k+1}$ is bigger than $x+y$

This is the strong form you speak of?

For all $n \in \mathbb{N}$

there is $c \in \mathbb{N}$ (so large that)

for all x, y coprime and bigger than c

$(R(x \cdot y \cdot (x+y)))^{1+1/n} \geq (x+y)$

$(R(x \cdot y \cdot (x+y)))^{(n+1)/n} \geq (x+y)$

$(R(x.y.(x+y)))^{n+1} \geq (x+y)^n$

$$(\forall \alpha)(\exists c)(\forall x, y > c)(R(x \cdot y \cdot (x+y))^{\alpha} \geq x+y))$$

$$(\forall k)(\exists c)(\forall xy > c)(R(x \cdot y \cdot (x+y))^{1/k} \geq x+y))$$

$$(\forall k)(\exists c)(\forall xy > c)(R(x \cdot y \cdot (x+y)) \geq (x+y)^k))$$

$$(\exists k)(\forall xy \text{coprime})((x+y) \leq (R(xy(x+y)))^k)$$

Presumably there are ternary versions of these conjectures?

The ABC conjecture alias the abc conjecture.

$sq(x)$ is the product of all the distinct prime factors of x.

$$(\forall \alpha)(\exists k)(\forall xy)(\frac{(sq(x \cdot y \cdot (x+y)))^{\alpha}}{x+y} \geq k)$$

Is the usual form. However, we can get rid of the reals easily enuff

$$(\forall \alpha)(\exists k)(\forall xy)((sq(x \cdot y \cdot (x+y)))^{\alpha} \geq k \cdot (x+y))$$

$$(\forall n)(\exists k)(\forall xy)((sq(x \cdot y \cdot (x+y)))^{1/n} \geq k \cdot (x+y))$$

$$(\forall n)(\exists k)(\forall xy)((sq(x \cdot y \cdot (x+y))) \geq k^n \cdot (x+y)^n)$$

Since k depends solely on n we can replace k^n by k getting

$$(\forall n)(\exists k)(\forall xy)((sq(x \cdot y \cdot (x+y))) \geq k \cdot (x+y)^n)$$

Actually we have to take k over to the other side, beco's k is really very much smaller than 1, so we replace it by $1/k$ and then take it over to the other side.

$$(\forall n)(\exists k)(\forall xy)((sq(x \cdot y \cdot (x+y))) \cdot k \geq (x+y)^n)$$

8.3 Agatha's theorem

Imre,

We start with the algebra $\mathcal{A}$ whose carrier set has no permutations with infinite support. (So $\mathcal{A}$ is amorphous. Admittedly that's not what i remember the conditions of the theorem as being, but i may have misremembered it)

We will use Ryll-Nardzewski.

We will find $g(k)$ depending only on k such that if there is an automorphism of $\mathcal{A}$ moving a k-tuple s to a k-tuple t then there is one that does it while moving no more than $g(k)$ things.

(Consider the $2k$-generator substructure $\mathcal{A}(s \cup t)$ generated by $s \cup t$. Any automorphism σ of $\mathcal{A}$ sending s to t must restrict to an automorphism of $\mathcal{A}(s \cup t)$. It would be nice if the permutation that is σ on $\mathcal{A}(s \cup t)$ and is the identity everywhere else is an automorphism of $\mathcal{A}$ did the trick, but i see no reason why it should, so we have to be subtle.)

Fix k and suppose there were no bound on the minimal size of the support of an automorphism sending one k-tuple to another. Then the set of ordered pairs of k-tuples from $\mathcal{A}$ would have a countably infinite partition, and i think that a bit of combinatorics will establish that if it has a countably infinite partition, then so does $\mathcal{A}$. But this needs to be checked. Let's call this upper bound $g(k)$, and march onwards.

(This matters beco's, for any concrete natural number (such as $g(k)$), quantifiers over automorphisms that move only $g(k)$ things are first-order quantifiers.)

Now consider the assertion that Symm($\mathcal{A}$), the automophism group of $\mathcal{A}$, has only 17 orbits on k-tuples. We can cook up a first-order formula as follows:

There are 17 k-tuples $s_1 \dots s_{17}$ such that for any other k-tuple t there is an automorphism moving only $g(k)$ things which moves t to one of the s_i.

This gives us a first-order axiom scheme which implies the oligomorphic condition that Ryll-Nardzewski talks about (Symm($\mathcal{A}$) has only finitely many orbits on k-tuples.)

All we need to do now is establish that $\mathcal{A}$ has the oligomorphic property. As you say, this is not obvious. However, we note that nothing done so far fully exploits the fact that $\mathcal{A}$ is an *algebra*. Presumably this is where we really need it.

(Brief reality check: why does this not work to show that all locally finite groups have countably categorical first order theory? Beco's $g(k)$ is not finite!)

8.4 souslin.tex

The reals have a countable dense subset. From this it follows that every set of disjoint open intervals is countable. Might there be a converse? In other words, must a dense total order of size (details!!) without an uncountable set of disjoint open intervals be isomorphic to the reals? We say that such a space satisfies the **countable chain condition**. (Spaces are assumed to be Hausdorff) Souslin's hypothesis is that the answer is 'yes'.

We will show that the answer is independent of ZFC. (Well, we'll get at least part of the way)

Martin's axiom arose from an attempt to separate the combinatorial content of CH from the cardinal arithmétic content.

Recall that a space is *Baire* iff the intersection of countably many dense open sets is dense. Baire's theorem asserts that every compact space is Baire. CH obviously enables us to strengthen this to "the intersection of fewer than $2^{\aleph_0}$ dense open sets is dense". This might look like a candidate for a weaker version of CH, but actually it's equivalent to CH. To get something weaker we have to restrict the spaces for which this is made. Let MA be

> Let $\mathcal{X}$ be a compact Hausdorff space satisfying ccc. Then every intersection of fewer than $2^{\aleph_0}$ dense open sets is dense.

This is not the way ZF-istes usually express it. It can be phrased as a fact about boolean algebras. This is important beco's of the appearance later of boolean valued models for set theory.

First we need a deviant concept of antichain. We say p and q are **incompatible** if there is no $r \leq p$ and $r \leq q$. An **antichain** is now a set of pairwise incompatible elements. We will say a poset satisfies the **countable chain condition** iff every (nudge, nudge) antichain is countable.

We can topologise the domain of any poset by taking basis sets to be $\{p : p \leq q\}$ for each q in the domain. The regular open algebra of this poset is a boolean algebra.

MA is now equivalent to the assertion that if $\mathcal{P}$ is a ccc poset and $D \subseteq \mathcal{P}(P$ with $|D| < 2^{\aleph_0}$. Then there is a "D-generic" filter [*HOLE definition in Dales and Woodin not consistent*]

We will now deduce SH from MA + $\neg$CH. Let $\langle P, <\rangle$ be a Souslin line. Let M be a maximal family of pairwise disjoint open intervals. Sine P is ccc, M is countable, so $\bigcup M$ is seperable. But P is not separable, so $\bigcup M$ is not dense in P, so there is an interval J disjoint from $\bigcup M$. Now $\langle J, <\rangle$ is a Souslin line too, so wlog we could have started with a Souslin line P whose every interval is nonseperable. Let us suppose we did this. So every open subset of P is a union of countably many pairwise disjoint nonseperable open intervals. Let us take these to be its **components**.

We will now construct a sequence $\langle U_\alpha : \alpha < \omega_1\rangle$ of dense open subsets of P s.t. for all $\alpha < \beta < \omega_1$,

1. $U_\beta \subset U_\alpha$
2. if I is a component of U_α, $\not\subset U_\beta$

Given U_α, we form U_β by deleting one point from each component of U_α. Since no component of U_α is seperable, the deleted point was not isolated in U_α [*HOLE why?*] so $U_{\alpha+1}$ is dense. (It's open beco's it's a union of things that are open-sets-minus-a-singleton – but such things are open)

Now for the limit case, λ.

For each $\alpha < \lambda$ we can find a countable S_α such that if $a < b$ belong to different components of U_α then there is $c \in S_\alpha$ with $a < c < b$. [*HOLE why?*] Let $S =: \bigcup\{S_\alpha : \alpha < \lambda\}$. Then set U_β to be the union of those intervals (a, b) which for all $\alpha < \lambda$ are included in a component of U_α.

To show U_λ is dense, let J be an open interval. J isn't seperable so there is an interval $(x, y) \subset J$ which meets S [*HOLE why?*] and $(x, y) \subseteq U_\lambda$ so U_λ meets J.

For $\alpha < \omega_1$ let Q_α be the family of components of U_α, and Q be the union of all the Q_α. Set $D_\alpha =: \bigcup\{Q_\beta : \alpha < \beta\}$. Now consider the poset of Q under subset. We claim this is ccc, and each D_α is dense in it. By MA there is a filter G in (*sic*) Q which meets every D_α. For each $\beta < \omega_1$ pick $J_\beta \in Q$ so that $G \cap Q_\beta = \{J_\beta\}$. Then for $\beta < \alpha < \omega_1$, J_α is a strict subset of J_β. So for every $\beta < \omega_1$ there is an open interval $U_\beta \subseteq J_\beta \setminus J_{\beta+1}$ which makes the U_βs an uncountable antichain, which is impossible.

DW p 100

H I A T U S

(liften from Devlin p 44)

A Souslin tree is a tree of size $\aleph_1$ all of whose branches are of length $< \omega_1$, with no uncountable antichains. SH iff there is no Souslin tree. One direction is easy: if there is a Souslin tree we can take

(Does the single worder have to ensure that for each node the set of its children (each "litter") is of order type omega?)

8.5 Fraenkel's conjecture

This is the result of a conversation with Robert Waters

The conjecture is that if $\mathcal{X} \subset \mathcal{P}(X)$ is closed under (binary) unions, and X is finite, then there is $x \in X$ belonging to at least half the elements of $\mathcal{X}$.

If $\mathcal{X}$ contains a singleton, $\{a\}$, say, then a clearly belongs to at least half the elements of $\mathcal{X}$. What if it contains a doubleton $\{a, b\}$? Well, $\mathcal{X}$ is the union of four sets, $A = \{X' \in \mathcal{X} : a \in X' \wedge b \notin X'\}$, $B = \{X' \in \mathcal{X} : b \in X' \wedge a \notin X'\}$, $AB = \{X' \in \mathcal{X} : a \in X' \wedge b \in X'\}$, and $E = \{X' \in \mathcal{X} : a \notin X' \wedge b \notin X'\}$.

We will be satisfied if one of the following hold:

$$|AB| + |A| \geq |B| + |E|$$

or

$$|AB| + |B| \geq |A| + |E|.$$

We know that $|AB| \geq |E|$ by closure under binary union, so it will suffice to show $|A| \geq |B| \vee |B| \geq |A|$. But this is immediate.

How about assuming that $\mathcal{X}$ contains a triplet, $\{a, b, c\}$. Then, analogously we seek one of

$$|ABC| + |AB| + |AC| + |A| \geq |BC| + |B| + |C| + |E|$$

$$|ABC| + |AB| + |BC| + |B| \geq |AC| + |A| + |C| + |E|$$

$$|ABC| + |AC| + |BC| + |C| \geq |AB| + |A| + |B| + |E|$$

As before $|ABC| \geq |E|$ so it would suffice to show:

$$|AB| + |AC| + |A| \geq |BC| + |B| + |C|$$
$$|AB| + |BC| + |B| \geq |AC| + |A| + |C|$$
$$|AC| + |BC| + |C| \geq |AB| + |A| + |B|$$

If these all fail we have

$$|AB| + |AC| + |A| < |BC| + |B| + |C|$$
$$|AB| + |BC| + |B| < |AC| + |A| + |C|$$
$$|AC| + |BC| + |C| < |AB| + |A| + |B|$$

Add all the LHS and the RHS and delete bilateral occurrences to get

$$|AB| + |AC| + |BC| < |A| + |B| + |C|$$

8.6 markstrom.txt

We have an infinite set C of chaps. For the moment assume that it is countable, so that without loss of generality each chap is a natural number. There is a strategy σ available to the chaps, with the property that, for any (infinite) subset $A \subseteq C$ (an allocation of hats to the chaps), and every $c \in C$, σ accepts as input the partition $\{A \setminus \{c\},\ C \setminus A \setminus \{c\}\}$ of $C \setminus \{c\}$ into two pieces (the allocation of hats to chaps other than c, and tells c what colour his hat is.

After consulting σ in this way, c puts on a hat of the colour σ told him his hat was. This gives us a new infinite subset of C, (a new allocation of hats) which we will call $\sigma(A)$. Thus σ defines a map from (infinite) subsets of C to subsets of C.

Now we are told that σ has this magical property: A and $\sigma(A)$ have finite difference!!

The challenge is to use this to devise a choice function on $\mathcal{P}(C)/Fin$.

Consider an equivalence class E of infinite subsets of C under finite difference. We want to use σ to pick a representative from E. By stipulation we know that σ sends E into itself. Now consider

$$\{n \in \mathbb{N} : (\exists e \in E)(|e \Delta \sigma(e)| = n)\}$$

This is a set of natural numbers, and so must have a least member, n_0, say. For all i know n_0 could be 0!). So let us consider

$$\{e \in E : |e \Delta \sigma(e)| = n_0\}$$

and it is in this set that we hope to find a representative for E.

ugh...

8.7 Joe Hurd on elliptic curve cryptography

See his home page for file of the same name.

Doubly periodic means that the function has two periods not that it is of arity 2 and is periodic in both arguments.

What is the connection with ellipses? What are the two periods?

Consider the ternary relation of "lie on a straight line"; It's functional beco's a cubic has precisely three roots.. And if the two arguments to the function are rational (resp. real) the value must be rational (resp. real) beco's a cubic cannot have precisely two rational (resp real) roots.

There is a unique field of power p^n, p a prime, known as $GF(p^n)$, and there are no fields of other powers.

Apparently for cryptographic purposes those that matter are $GF(p)$, p a large prime, or $GF(2^n)$

Projective space of dimension 1 over $\mathbb{R}$ is the set of straight lines through the origin in $\mathbb{R}^2$. Or the unit circle with antipodal points identified. Better thought of as a line and a point, giving rise to the set of lines through the point plus the parallel line which is the point at infinity.

Projective space of dim 2 over $\mathbb{R}$ is the set of lines through the origin in $\mathbb{R}^3$.

Reflect a point in the x-axis is a kind of negation. $A + B$ is the reflection in the x axis of the third point lying on the line AB. To obtain $A + A$ you use the tangent at A. This is called the **tangent-chord method**. Why is it associative?

If an equation is homogeneous (every term is of the same degree) then the set of solutions forms a vector space.

Discrete logarithm problem. Given $x, y \in G$ find $k \in \mathbb{N}$ s.t. $x^k = y$. This is hard if G is the multiplicative group of the integers mod p, but still subexponential: $e^{\log(n)/3}$. It seems to be harder using elliptic curve groups over finite fields.

Bob generates $g^x = h$ and publishes g and h but does not publish x.

Alice, using g and h, generates a number k (less than the size of the group) computes $a = g^k$ and $b = h^k \cdot m$ and sends $\langle a, b \rangle$ to Bob.

Bibliography

[1] Randall Holmes "The universal separable metric space of Urysohn and isometric embeddings thereof in Banach spaces" FM **140** (1992)