Notes on (pretty much) all of mathematics

Krzysztof Szymański
338068@uwr.edu.pl

March 10, 2025

Contents

Chapter 1

Naive set theory

Lemma 1.1. *Let*

$$f : X \to X$$

be a function from a finite *nonempty set into itself. Then, you can uniquely decompose the set X into disjoint parts*

$$X = X_1 \cup X_2$$

such that f restricted to X_1 is a permutation and every element from X_2 eventually gets to X_1, that is

$$\forall x_2 \in X_2 \exists k\, f^k \in X_1.$$

Proof. Define a decreasing (TODO: why) sequence of sets by

$$X^0 = 0, X^{k+1} = f[X^k].$$

Every set in the sequence is finite and nonempty. Since the sequence is decreasing, it must stabilize eventually. Let k_0 be the first index at which it stabilizes, i.e.

$$X^k = X^{k+1} = f[X^k].$$

Then

$$f : X^k \to X^k$$

is surjective, so by a previous lemma it is bijective. We take

$$X_1 = X^k,\, X_2 = X \setminus X^k$$

to form the desired decomposition.

□

Visualization If we want to visualize how such a function looks, let us first recall a visualization of permutations – they form cycles! The elemnts outside of the *permutation set* X_1 form trees, whose roots are vertices of X_1.

Example 1.1. *Consider the function …*

This may seem like a silly little fact, but the ideas come up often: in the Lefschetz Fixpoint theorem, Jordan decomposition from linear algebra …

How do we extend this to infinite sets? Let's try using a fixpoint theorem! In this case, we'll be looking for the largest fixpoint of a decreasing function. Note that on infinite sets the trees mentioned earlier may be infinite (consider the example of $x \mapsto x+1$ on $\mathbb{N}$). But these are actually permutations! So maybe the structure looks the same? But how do we define the *permutation set*?

Attempt 1: take the maximal set from

$$\{A \mid f \text{ is a permutation on } A\}$$

§1.1 Literature used for this chpater

Chapter 2

Point-set topology

This chapter summaries basic, point-set topology. It is roughly equivalent to Chapter 2 of Munkres, without the metric space material.

§2.1 The location trilemma

Fix a topological space X, a set A an a point x. In terms of set theory, a point has two *locations* with respect to A – either $x \in A$ or $x \notin A$. In topology, we care about a little more. Although any endpoint of a closed interval $[a, b]$ certainly is *in* the interval, we would not say it is *inside* that interval. The difference becomes even starker when looking at sets like $\mathbb{N} \subseteq \mathbb{R}$ or $\mathbb{Q} \subseteq \mathbb{R}$.

The right way – or at least *a right way* – is to say that a point x is *inside* a set A, if apart from $x \in A$, a whole neighbourhood $U \ni x$ is contained in A. Since this readily implies $x \in A$, we may dispense with that requirement in any definition. Metaphorically, this means that x *cannot escape* A or *everything* x *sees is inside of* A.

The notion that $x \notin A$ has a very similar topological analogue. We will say that x *is outside of* A if not only $x \notin A$, but a whole neighbourhood $U \ni x$ is disjoint from A. Another word for that is that x *is separated from* A. This also provides the metaphorical meaning.

With these two notions lifted up from mere set-based combinatoris to topology, we have made a bit of a problem. It could be the case that for a point x and set A neither of the above properties hold. Pictorially speaking, x is then *not far from* A, but also not really *inside of* A – a kind of in-between state. In this case we call x a *boundary point of* A.

Topological properties of operators (such as Int, Cl, Bd etc.) can often be deduced from this simple observation, which we state as a lemma. This was first taught to me by Krzysztof Omiljanowski. I am not aware of any name for the fact, so I made up my own.

Lemma 2.1. *(The location trilemma.) Let X be a topological space, $A \subseteq X$ and $x \in X$ (not neccesarily $x \in A$). Then exactly one of the following is true:*

1. *For some neighbourhood U, $x \in U \subseteq A$. Then x is called an* interior point*.*
2. *For some neighbourhood U, $x \in U \subseteq A^c$. Then x is called an* exterior point*. We also say that x is* separated from *A by the neighbourhood U.*
3. *For all neighbourhoods U, $U \cap A \neq \varnothing$ and $U \cap A^c \neq \varnothing$. Then x is called a* boundary point*.*

Proof. Clause (1) implies that $x \in A$ and (2) implies that $x \in A^c$, so they cannot hold at once. Moreover both of them imply that (3) is false. We will show that if both (1) and (2) do not hold, then (3) does.

Pick any neighbourhood $U \ni x$. Since (1) does not hold, $U \nsubseteq A$, so $U \cap A^c \neq \varnothing$. Analogously, $U \cap A \neq \varnothing$. □

The names we gave to the properties of points with respect to A are not random – they correspond exactly to the usual definitions of Int, Bd, Ext.

§2.2 The closure axiomatization

As one should know, a topology on X might as well be defined in terms of which sets are closed rather than which sets are open. In practice, another way of looking at closed sets might also pop up. In many scenarios, there is an operator Cl, which we might call a *closure* operator, of the signature

$$\mathrm{Cl} : \mathcal{P}(X) \to \mathcal{P}(X),$$

which adjoins to a set A some elements that are in a given sense *reachable*, *deducible*, *obtainable* etc. from A.

An example of such an operator would be, for any given topology $\mathcal{T}$ on X, the closure operator of that topology. One might wonder if from an operator one might recover a topology. If we are going to do that, there are a few questions we need to answer.

How do we recover open sets? We can get open sets as complements of closed sets. Then the question is how do we recover closed sets. The key property to use is that a set C is closed iff $\mathrm{Cl}\, C = C$.

Definition 2.1. *An operator*

$$\mathbf{c} : \mathcal{P}(X) \to \mathcal{P}(X)$$

is said to satisfy the Kuratowski closure axioms if it satisfies to following

TODO: add this to Anki, make a flashcard, whatever

(K1) it preserves $\varnothing$, i.e. $\mathbf{c}(\varnothing) = \varnothing$;

(K2) it is extensive*, i.e. $A \subseteq \mathbf{c}(A)$ for all A;*

(K3) it is idempotent*, i.e. $\mathbf{c}(\mathbf{c}(A)) = \mathbf{c}(A)$ for all A;*

(K4) it distributes over finite sums, i.e. $\mathbf{c}(A \cup B) = \mathbf{c}(A) \cup \mathbf{c}(B)$ for all A, B.

Now we prove that this actually defines a topology.

Lemma 2.2. *Let $\mathbf{c}$ be an operator*

$$\mathbf{c} : \mathcal{P}(X) \to \mathcal{P}(X)$$

satisfying the Kuratowski closure axioms. Then, the collection of its fixpoints, i.e.

$$co\mathcal{T} = \{A \mid \mathbf{c}(A) = A\}$$

defines a topology as its family of closed sets.

Proof. From the axiom (K1) we get that $\varnothing \in co\mathcal{T}_{\mathbf{c}}$. We also need $X \in co\mathcal{T}$, which follows from (K2), as no set in X can be larger than all of X. Suppose now that A, B are closed in the sens above. Then we have

$$\mathbf{c}(A \cup B) = \mathbf{c}(A) \cup \mathbf{c}(B) = A \cup B,$$

so $A \cup B$ is closed too. The intersection property is the tricky part.

Take $C_i \in co\mathcal{T}$. Then

$$\bigcap_i C_i \subseteq C_j$$

for all j, so

$$\forall j.\, \mathbf{c}\left(\bigcap_i C_i\right) \subseteq \mathbf{c}(C_j) = C_j,$$

so

$$\mathbf{c}\left(\bigcap_i C_i\right) \subseteq \bigcap_j C_j.$$

The other inclusion follows by extensivity of $\mathbf{c}$, so we actually have the equality

$$\mathbf{c}\left(\bigcap_i C_i\right) = \bigcap_j C_j.$$

□

You may have noticed that (K1), (K2) and (K4) were used in the proof, but not (K3). Then an operator satisying all of the Kuratowski axioms except for (K3) defines a topology via closed sets. On the other hand, the closure operator of that topology definitely has property (K3), so these two are different operators!

Luckily, we can recover the closure operator using lattice theory. First, a definition and a lemma.

Definition 2.2. *An operator*

$$\mathbf{c} : \mathcal{P}(X) \to \mathcal{P}(X)$$

is called a Čech closure operator *if it satisfies the Kuratowski axioms (K1), (K2) and (K4).*

Lemma 2.3. *Any operator satisfying (K4) is* monotonic*, i.e.*

$$A \subseteq B \Rightarrow \mathbf{c}(A) \subseteq \mathbf{c}(B).$$

for all $A, B \subseteq X$.

Proof. TODO! □

Now that we've seen these, we'll want to recover the actual A. If you think of the operator $\mathbf{c}$ as *enriching* a set, a closed set is one that *is completely enriched*. Then, how do we get $\operatorname{Cl} A$? Well, we start by enriching it

$$A \to \operatorname{Cl} A,$$

but this may not be enough, so we enrich the result and get

$$A \to \operatorname{Cl} A \to \operatorname{Cl}(\operatorname{Cl} A),$$

but this may not be enough, so we enrich the result

$$A \to \operatorname{Cl} A \to \operatorname{Cl}(\operatorname{Cl} A) \to \operatorname{Cl}(\operatorname{Cl}(\operatorname{Cl} A)),$$

and... we're stuck in an infinite loop. In such cases, it helps to take a peek past infinity and consider

$$\bigcup_{k=0}^{n} \operatorname{Cl}^k A.$$

This still does not work.

Example 2.1. *You can define a Cech closure operator which fails this. Let*

$$X = \mathbb{N} \cup \{\infty\}$$

and

$$\mathbf{c}(A) = \begin{cases} \varnothing & \text{when } A = \varnothing \\ X & \text{when } \infty \in A \\ \{0, 1, \ldots, \sup A + 1\} & \text{when } A \neq \varnothing \text{ and } \infty \notin A \end{cases}$$

You can check that this is a Cech closure operator, but the only closed sets in $\mathbf{co}\mathcal{T}_{\mathbf{c}}$ *are* $\varnothing$ *and* X*, i.e. this is the indiscrete topology on* X*.*

In reality you need a transfinite number of iterations. However, for a Kuratowski operator this chain stabilizes almost immediately, so we're saved!

Lemma 2.4. *For a Kuratowski closure operator* $\mathbf{c}$*, the closure operator of its generated topology* $\mathcal{T}_{\mathbf{c}}$ *is precisely* $\mathbf{c}$*.*

Proof. Let $C \supseteq A$ be the topological closure of A. Then

$$\mathbf{c}(A) \subseteq \mathbf{c}(C) = C.$$

Since $\mathbf{c}$ is idempotent by (K3), $\mathbf{c}(A)$ is closed in $\mathcal{T}_{\mathbf{c}}$. Since a topological closure is the smallest closed set containing A, we have that

$$C \subseteq \mathbf{c}(A) \Rightarrow C = \mathbf{c}(A)$$

$\square$

Connection with lattice theory. Note that we want to find the least (because topological closure is the smallest closed set) fixpoint (because of the definition of topology) of $\mathbf{c}$ greater than A. This is exactly the setting of the Kleene fixpoint theorem, and what we're doing here is using that.

§2.3 Literature used for this chapter

1. Munkres, chapter 2
2. Wikipedia, Kuratowski Closure Axioms
3. Wikipedia, Preclosure operator

Chapter 3

Abelian groups

We will now explore some features of abelian groups. These will later be recontextualized into features of modules (for those of you know, an abelian group is essentialy the same as a $\mathbb{Z}$-module). In this chapter, we will derive a structure theorem.

§3.1 Basic properties

§3.2 Torsion

Central to the study of abelian groups is the notion of torsion. We begin with two definitions, one for elements and one for groups.

Definition 3.1. *An element of an abelian group is called a* ***torsion element*** *if it is of finite order. An element of infinite order is called* ***torsion-free****.*

Definition 3.2. *An abelian group is called a* ***torsion group*** *if all its elements are torsion. An abelian gropu is called* ***torsion-free*** *if all its* nonzero *elements are torsion-free.*

One might ask why we define this only for abelian groups. The following crucial fact is the reason – more precisely, the fact that its proof relies on commutativity.

Theorem 3.1. *If α, β are two torsion elements then their sum (and difference) is also a torsion element. Thus, the torsion elements form a subgroup.*

Proof. Let α have order n and β have order m. Then we have

$$nm(\alpha + \beta) = nm\alpha + nm\beta = m(n\alpha) + n(m\beta) = m \cdot 0 + n \cdot 0 = 0.$$

The subgroup claim is proved by noting that the same computation hold as well for difference and that the identity element is of course torsion. □

Since we have a well defined kind of subgroup, we might as well give it a name.

Definition 3.3. *Let G be an abelian group. Its* ***torsion subgroup****, denoted*

$$\mathrm{T}(G)\mathrm{or}Tor(G)$$

is the subgroup consisting of all the torsion elements of G.

We may hope that such a nice notion of substructure is also valid for the torsion-free elements of a group. This fails, however, as the set of torsionfree elements is the complement of a subgroup, which will very often fail to be a subgroup. On the level of elements, we have the following fact.

Theorem 3.2. *Let $a \in G$ be torsionfree and $b \in Tor(G)$. Then $a + b$ is torsionfree.*

Proof. We give two proofs. For the first one, suppose $a + b$ is torsion. Then

$$a = (a + b) - b$$

would be torsion (since the torsion elements form a subgroup).

For the second proof, suppose $a + b$ is torsion with order n and that b has order m. Then

$$0 = mn(a + b) = mna + mnb = mna,$$

so a is torsion as well. □

All is not lost though! We cannot define a notion of *torsion-free subgroup*, but there is a simple way of killing the torsion – quotienting!

3.2.1 A weird example of groups

3.2.2 Functors

3.2.3 Tensors

§3.3 Structure theorem

§3.4 A smudge of infinite abelian group theory: the prufer groups

Prufer groups are: divisible, p-torsion, subgroups, infinitely generated

§3.5 Sources for this chapter

1. Ludomir Newelski, Algebra II
2. Ludomir Newelski, Algebra 2R
3. Wikipedia: torsion subgroup, torsion, torsion-free group, Prufer groups

Chapter 4

The Künneth formula

This chapter accomplishes a lofty goal: to calculate the homology groups of a product topological space.

§4.1 Setup

We will first try and discover how this all works for CW-homologies, and then use what we've found to generalize the statements to singular homology. The reason why CW-complexes are good for the job is that given CW-structures on X and Y, we can easily form a CW-structure on $X \times Y$. One should expect to form a simplicial structure on a product of simplicial complexes, but that would require subdividing a product of n- and m-simplices into $(n+m)$-simplices, which, if anything, is unpleasant.

We have a product structure on $X \times Y$ given by products of cells and products of characteristic maps. Let us now try to calculate the cellular boundaries.

One of the cells is larger in dimension: then we get a coefficient of zero, because the map turns out to be constant.

If the above does not hold, then we have to reduce exactly one dimension by one. However, if in the dimension we keep we use a different cell of that dimension, the same problem applies.

Generally, we have the following:

Lemma 4.1. *Suppose we have a continuous map*

$$f : I^k \to I^{k-1},$$

which is equal to a product map

$$f = id_I \times g$$

for some continuous

$$g : I^{k-1} \to I^{k-2}.$$

Then, the degree of

$$\widetilde{f} : \operatorname{Bd} I^k \to I^{k-1} / \operatorname{Bd} I^{k-1}$$

is equal to the degree of

$$\widetilde{g} : \operatorname{Bd} I^{k-1} \to I^{k-2} / \operatorname{Bd} I^{k-2}.$$

Remark. Note that since this is not the same sphere, the degree only makes sense *modulo orientation*.

§4.2 Orientations

Remember that when describing a CW-complex structure, there is an additional degree of freedom – the orientation of each cell. This orientation needs to be taken into account.

For the purposes of this chapter, it is very important to see how an orientation of the interior can be turned into an orientation of the boundary.

One way of doing this would be to just take the *vector to the right and up* in an Euclidean space. However, this leads to an inconsistent choice of orientation on the boundary. A correct way is to pick an inward facing normal at each point on the boundary and then pick an orientation on the boundary which

This can formally be described in homology by noting that the boundary homomorphism is an isomorphism. The determined choice of generator for I^k then projects down to $\operatorname{Bd} I^k$.

§4.3 Literature used for this chapter

1. Allen Hatcher, Algebraic Topology
2. MathOverflow, biography of Hermann Kunneth: https://mathoverflow.net/questions/114215/who-was-hermann-k
3. https://sili-math.github.io/AT2020/Lecture-22.pdf

Chapter 5

Local properties in algebra

§5.1 Localization of modules

§5.2 Local rings

This section is devoted to a class of rings very important in algebraic geometry – local rings. Although the definition is purely algebraic, it corresponds to notions of *looking around a point* on a (affine) variety.

5.2.1 Characterisations of local rings

All material in this section applies to (unital) rings. They need not be commutative for the results to work. All the proofs in this section are my own work.

Definition 5.1 (local ring)**.** *A ring R having a unique maximal ideal $\mathfrak{m}$ is called a* **local ring** *and denoted $(R, \mathfrak{m})$.*

It turns out there is another characterisation of such rings, which is of particular usefulness in working with the second point of interest of this chapter – Discrete Valuation Rigs, whose definition we postpone for now.

Theorem 5.1 (local ring characterisation)**.** *A ring R is local iff the set of noninvertible elements forms an ideal. This is the the unique maximal ideal.*

Proof. The first implication is trivial, since an ideal containing a unit is not proper.

Suppose now R is local with maximal ideal $\mathfrak{m}$. Since $\mathfrak{m}$ is proper, it does not contain any invertible element. Take a noninvertible $r \in R$. The principal ideal (r) can be extended (via Zorn's lemma) to a maximal ideal, which must be the unique maximal ideal $\mathfrak{m}$, so $r \in \mathfrak{m}$.

□

There is also an element-wise characterisation of local rings. To see it emerge, let us try a different line of attack for the previous proof.

We know that $\mathfrak{m}$ contains only noninvertible elements, but we still have to prove that it contains all such elements. By way of contradiction, suppose $r \notin \mathfrak{m}$. Now by maximality we have

$$\mathfrak{m} + (r) = R$$

and in particular

$$m + ar = 1$$

for some $a \in R, m \in \mathfrak{m}$. We would like to derive a contradiction, so we would hope that the identity above lets us conclude that r is invertible. The property below is what we need

Theorem 5.2 (local ring addition). *A ring R is local iff it has the following property: for all a, b such that*

$$a + b \in R^\star$$

at least one of a, b is invertible.

Proof. The line of attack above shows how this implies R being local. Once we have this property, we can take any maximal ideal. If there was some noninvertible element it did not contain, we would be able to extend it.

On the other hand, take a local ring. We will show it is impossible to sum two noninvertibles to a unit. We have that

$$(a) + (b) \subseteq \mathfrak{m} + \mathfrak{m} = \mathfrak{m},$$

which contains no units.

□

It should be noted that the property expressed in the previous theorem might be rephrased in two ways: first, take an arbitrary finite sum instead of two elements and second, have the sum be equal to 1 instead of invertible. It should be easy to see that all such characterisations are equivalent.

5.2.2 Examples and generic constructions

5.2.3 Properties of local rings

Krull's intersection, Nakayama's, Kaplansky's theorem

§5.3 Valuations on Fields and Discrete Valuation Rings

§5.4 Literature used for this chapter

1. Piotr Kowalski, Algebraic Curves. Chapter 2.3.
2. nlab page for local rings
3. https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=2933&context=td (for Kaplansky's theorem)
4. Wikipedia pages

Chapter 6

Algebraic Geometry, problemset 5

Problem 4

Since v is a smooth point of C, the ring $\mathcal{O}_v$ is a DVR. A local parameter is defined as any uniformizing parameter of that DVR and having a zero of order 1 means being of valuation 1 in that DVR.

Thus is suffices to prove the following: in a DVR, an element is of valuation 1 iff it is irreducible.

Since a DVR is a local ring, let us denote its maximal ideal by $\mathfrak{m} = (r)$. Note that r must be irreducible, otherwise the ideal $\mathfrak{m}$ would not be maximal. The valuation on this DVR can be described as the usual r-adic valuation or belonging to a power of the maximal ideal (see solution of **Problem 7.**).

Let f be a local parameter, i.e. irreducible. Then, by **Remark 2.37.**

$$(f) = \mathfrak{m} = (r),$$

so f and r are associates, therefore $f = ur$ for some unit u. We have that $u \in \mathcal{O}_v \setminus \mathfrak{m}$, so $v(u) = 0$ and

$$v(f) = v(u) + v(r) = 0 + 1 = 1.$$

On the other hand, let f be of valuation 1. Then, as the valuation is the r-adic valuation we have that

$$f = ra$$

for some $r \nmid a$, so $a \in \mathcal{O}_v \setminus \mathfrak{m}$. This implies a is a unit, so f and r are associated, so f is irreducible since r is.

Problem 5

Problem 5a

Consider a tangent line $L = V(\alpha X + \beta Y + \gamma)$.

If L is tangent, the intersection number $I(0, L \cap C) > 1 > 0$, so $0 \in L$, so $\gamma = 0$. By the same reasoning, F has zero constant term.

It is easy to see that a line is a smooth curve. Therefore, the curve C is tangent to L iff F is of valuation at least two in $\mathcal{O}_0$ (which is a DVR). By **Problem 7** that is equivalent to F being a member of $\mathfrak{m}_v^2$ (where F is reinterpreted as $F + I(L)$ in $K(L)$).

Algebraically, the square of the maximal ideal corresponds to

$$I_L^2(0)\mathcal{O}_0 = \left\{ \frac{G}{H} : G \in I_L^2(0), H \notin I_L(0) \right\}.$$

Thus, F being of valuation at least 2 is equivalent to it being of the form

$$F = \frac{1}{N} \cdot \sum_i G_i H_i$$

as a rational function in $K(L)$ for some $N \in K[L] \setminus I_L(0)$ and $G_i, H_i \in I_L(0)$. This is again equivalent to the identity

$$FN = \sum_i G_i H_i$$

in $K[L]$, which in turn is equivalent to F having the form

$$FN = \sum_i G_i H_i + P(\alpha X + \beta Y)$$

for some polynomials such that $N(0) \neq 0$ and $G_i(0) = H_i(0) = 0$.

After this introduction, we will show

$$T_0 C = V\left(\partial_X F(0) X + \partial_Y F(0) Y\right)$$

via two inclusions, first from left to right.

If both partials are 0, the vanishing set is the whole space $\mathbb{A}^2$, so this inclusion is trivial. Suppose now at least one partial is nonzero. Differentiating both sides of the above identity w.r.t. X we get that

$$\partial_X F \cdot N + F \cdot \partial_X N = \left(\sum_i \partial_X G_i \cdot H_i + G_i \cdot \partial_X H_i \right) + \partial_X P \cdot (\alpha X + \beta Y) + \alpha P.$$

Evaluating both sides at 0 and remembering which polynomials vanish at 0 we obtain

$$N(0)\partial_X F(0) = P(0)\alpha.$$

We can now repeat this for differentiation w.r.t. Y. This implies that the vectors

$$[\partial_X F(0), \partial_Y F(0)]^T, [\alpha, \beta]^T \in \mathbb{K}^2$$

are linearly dependent, so they describe the same line.

Now for the other inclusion. We will lean on the fact that the first partials evaluated at zero are the coefficients of the monomials X, Y in the polynomial F. Suppose first both partials of F are zero. Then we can write F as

$$F = X^2 G(X,Y) + XYH(X,Y) + Y^2 P(X,Y) + 0 \cdot (\alpha X + \beta Y)$$

which is the form we needed. This gives that *any* line is tangent to C at 0, so the tangent space is the whole plane (as is $V(0 \cdot X + 0 \cdot Y)$).

If, on the other hand, some partial is nonzero, we can write F as

$$F = 1 \cdot (\partial_X F(0) \cdot X + \partial_Y F(0) Y) + X^2 G(X,Y) + XYH(X,Y) + Y^2 P(X,Y).$$

This proves that C is indeed tangent to $V(\partial_X F(0) \cdot X + \partial_Y F(0) Y)$.

Problem 5b

From the previous subproblem and the lecture we know that both T_0C and $I_C(0)/I_C(0)^2$ are finite-dimensional K-vector spaces. This allows us to use a theorem of linear algebra which states that a bilinear map such as is given in the problem induces an isomorphism iff for all $P \neq 0$ the function

$$x \mapsto \Phi(x, P)$$

is nonzero (i.e. is nonzero for some x) or, equivalently, that if this function is zero then so is P. Here P should be understood as the representative regular function in $K[C]$ or a polynomial which represents that function (so a representative of representatives).

Suppose then that this function is zero. First consider the case that both partials of F are zero. Then the tangent space is the whole plane and we have that for all $x, y \in K$

$$\partial_X P(0)x + \partial_Y P(0)y = 0,$$

which gives that both partials of P vanish at zero, so we have the form

$$P = X^2P_1 + XYP_2 + Y^2P_3$$

for P as a polynomial, which in particular implies

$$P \in I_C(0)^2$$

as a regular function, so

$$P = 0$$

in $I_C(0)/I_C(0)^2$.

Now consider what happens when F has a nonzero partial. This means that

$$\partial_X P(0)x + \partial_Y P(0)y = 0$$

for all points such that

$$\partial_X F(0)x + \partial_Y F(0)y = 0,$$

which gives that either the gradient of P is zero (in which case we can repeat the reasoning from the previous case) or that the gradients of P and F are linearly dependent. Then there exists a scalar α such that

$$P - \alpha F$$

has zero first partials, so

$$P = \alpha F + X^2P_1 + XYP_2 + Y^2P_3$$

as a polynomial, so

$$P = X^2P_1 + XYP_2 + Y^2P_3 + I(C)$$

as a regular function. As we have done many times, we now conclude that

$$P \in I_C(0)^2,$$

which is what we needed.

Problem 6

Problem 6a

Let

$$a = r^n \frac{\alpha}{\beta}, b = r^m \frac{\gamma}{\delta},$$

with $r \nmid \alpha, \beta, \gamma, \delta$ and wlog $n \geqslant m$. Then

$$a + b = r^m \frac{r^{n-m}\alpha\delta + \beta\gamma}{\beta\delta}.$$

Note that r does not divide the numerator (since r is irreducible and thus prime in a UFD). If $n > m$, it will also not divide the denominator, but if $n = m$ it might. In the first case the valuation is exactly m, while in the second case it might change – but only increase.

Problem 6b

We have

$$ab = r^{n+m} \frac{\alpha\gamma}{\beta\delta}.$$

By virtue of r being prime, we have $r \nmid \alpha\gamma\beta\delta$, so

$$v_r(ab) = n + m.$$

Problem 6c

For every $n \in \mathbb{Z}$ we have

$$v_r(r^n) = n,$$

so the valuation is indeed surjective.

Problem 7

We claim that

$$\mathfrak{m}^n = \{x : v_R(x) \geqslant n\},$$

from which the problem follows immediately. For $n = 0$ the claim follows from the nonnegativity of the valuation.

Let $r \in R$ be the uniformizing parameter. Then we have that

$$\mathfrak{m} = (r)$$

by **remark 2.37.**, so an element of $a \in \mathfrak{m}^n$ is of the form

$$a = \sum_i a_i r^n$$

for some $a_i \in R$, so

$$v_R(a) \geqslant \min(v_r(a_1 r^n), v_r(a_2 r^n), \ldots, v_r(a_j r^n)) \geqslant \min(n, n, \ldots n) = n.$$

This gives us one inclusion of the claim. For the other one, take an element x of valuation no less than n. The definition of valuation implies that for some $m \geqslant n$

$$x = r^m y = (r^{m-n} y) r^n \in \mathfrak{m}^n.$$

This completes the solution.

Problem 8

Take $x, y \in \mathcal{O}_v$. Then we have

$$v(xy) = v(x) + v(y) \geqslant v(y) \geqslant 0$$

and

$$v(x+y) \geqslant \min(v(x), v(y)) \geqslant 0.$$

This implies that $\mathcal{O}_v$ is a subring of L. If we take x, y such that both valuations are positive, the sum has a positive valuation, and if at least y has a positive valuation then the product does as well. This implies that $\mathfrak{m}_v$ is an ideal.

Note that for any valuation we have

$$v(1) = v(1 \cdot 1) = v(1) + v(1),$$

so

$$v(1) = 0.$$

Now take any $x \in \mathcal{O}_v \setminus \mathfrak{m}_v$. This means that $v(x) = 0$. Let y be the multiplicative inverse (in L!) of x. Then

$$0 = v(1) = v(xy) = v(x) + v(y),$$

so $v(y) = 0$ and $y \in \mathcal{O}_v$. If $v(x) > 0$, then $v(y) < 0$ and $y \notin \mathcal{O}_v$. We have just proved that the set of noninvertible elements of $\mathcal{O}_v$ is an ideal. Therefore $(\mathcal{O}_v, \mathfrak{m}_v)$ is a local ring.

Since the valuation is surjective, $\mathfrak{m}_v$ is nonempty and $\mathcal{O}_v$ has nonzero noninvertible elements, so it is not a field. To finish the proof that $\mathcal{O}_v$ is a DVR all we need to do is show that $\mathcal{O}_v$ is a PID.

To do that, take r to be any element of valuation 1. Such an element exists by surjectivity of the valuation.

Claim. Let $t \in \mathcal{O}_v$ and $v(t) = n \geqslant 0$. Then $t = ur^n$ for some unit $u \in \mathcal{O}_v$.

Proof. Let

$$\alpha = \frac{t}{r^n} \in L.$$

Then α has valuation 0, so it actually is an invertible element of $\mathcal{O}_v$.

Now take an ideal I. Note that if I has any element of valuation n, then by the claim above it contains all elements of valuation n as they all are associated with r^n. It will also contain an element (so all elements) of higher valuations by virtue of being closed under multiplication by r.

This lets us conclude that

$$I = (r^n)$$

where n is the smallest valuation achievable by an element of I.

Chapter 7

Approximation properties of smooth functions

Smoothness is a really good property that a function can have. However, a lot of the functions we need aren't smooth, i.e. ReLU, $\max(0, \min(1, x))$, $\chi_{[a,b]}$ and so on. In this chapter, we will show how to effectively approximate such functions with very good smooth lookalikes.

§7.1 Characteristic functions on intervals

Let us make our way towards approximating the interval characteristic function. In the special case of the real line, the story begins with quite a wonderful function – the magnificent

$$e^{-1/x^2}.$$

Taken literally, it's not quite defined at 0. However, as $x \to 0$, $-1/x^2 \to -\infty$, so we can define the value at 0 to be 0. Notice that the convergence of the argument to $-\infty$ is really fast – fast enough the we can retain smoothness at 0. Let us state the properties formally.

Lemma 7.1. *The function*

$$g(x) = \begin{cases} e^{-1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is smooth and satisfies the following properties:

1. $g(0) = 0$,
2. $g(x) \leq 1$,
3. $\lim_{x\to 0} g(x)/P(x) = 0$ *for any* $P \in \mathbb{R}[X]$,
4. $g^{(n)}(0) = 0$ *for all* n.

Proof. Properties (1) and (2) should be obvious. For (3) we perform the change of variables $u := 1/x^2$ to obtain

$$\lim_{x\to 0} \frac{g(x)}{P(x)} = \lim_{u\to\infty} \frac{e^{-u}}{P(\pm 1/\sqrt{u})} = \lim_{u\to\infty} \frac{u^{\deg P/2} e^{-u}}{Q(\pm\sqrt{u})} = 0,$$

where $Q \in \mathbb{R}[X]$ (actually, it's P with the coefficients written backwards). We use (3) to derive an (4). By induction, we have that at all $x \neq 0$

$$g^{(n)} = \frac{g(x)P_n(x)}{Q_n(x)}$$

for some $P, Q \in \mathbb{R}[X]$ and the only root of Q is 0. Indeed,

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{g(x)P_n(x)}{Q_n(x)} = \frac{-\frac{1}{x^2}Q_n(x)P_n(x)g(x) + P_n'(x)Q_n(x)g(x) - P_nQ_n'g(x)}{Q_n^2(x)} = g(x)\frac{P_{n+1}(x)}{Q_{n+1}(x)},$$

where we have expanded the fraction by x^2 to get the desired form. This gives that

$$Q_n = x^{a_n},$$

where

$$a_0 = 0$$
$$a_{n+1} = 2a_n + 2,$$

from which we can derive

$$a_n = 2(2^n - 1).$$

□

The magic of this function is that, since it has all the derivatives at 0 equal to 0, we might as well say it's zero on all negative numbers *and still get a smooth function*! This is very much not the case for more primitive attempts to do this – look at the case of x or more generally x^k and see that it only gives as a function of class C^{k-1}.

This maneover let us discard half of the real line a smooth fashion, and this means we're halfways towards approximating $\chi_{[0,1]}$. The idea will be to use functions like this:

$$g(x)g(1-x).$$

This has a small bump, but since g is very flat near 0, the product of two such functions will not be very large. The solution solution to this problem is to widen the interval for the moment, so that values closer to 1 get multiplied together in the formula above. Let us define

$$h_n = g(x)g(n-x).$$

This should look better, but now

$$\operatorname{supp} h_n = [0, n]$$

where we would much prefer

$$\operatorname{supp} h_n = [0, 1].$$

No worries – we can see that the *shape* is right, so squeezing the function will do.

Lemma 7.2. *The function sequence*

$$f_n(x) = h_n(nx) = g(nx)g(n - nx).$$

converges to $\chi_{[0,1]}$ *pointwise a.e. (outside of* $\{0, 1\}$*) and in* L^1.

Proof. Let us quantify the idea that f_n *is almost* 1 *on almost all of* $[0,1]$. In terms of h_n, these functions should be almost 1 on almost all of $[0,n]$. We will pick an α (varying with n), for which we will be able to establish a good bound. Suppose

$$\alpha n \leq x \leq (1-\alpha)n.$$

This gives

$$\begin{aligned}
\frac{1}{\alpha^2 n^2} &\geq \frac{1}{x^2} \\
\frac{2}{\alpha^2 n^2} &\geq \frac{1}{x^2} + \frac{1}{(n-x)^2} \\
-\frac{2}{\alpha^2 n^2} &\leq -\frac{1}{x^2} - \frac{1}{(n-x)^2} \\
e^{-\frac{2}{\alpha^2 n^2}} &\leq h_n(x)
\end{aligned}$$

We now want to pick a sequence α such that both

$$\lim_n \alpha = 0$$

and

$$\lim_n \frac{2}{\alpha^2 n^2} = 0.$$

Of course, $\alpha = 1/\sqrt{n}$ does the trick nicely. We can get better convergence estimates for a different choice. $\square$

Another kind of useful function we might like to have are piecewise linear functions. To effectively approximate those, we need just see how to approximate ReLU and ... – all piecewise linear functions are sums of shifted/scaled combinations of these two.

§7.2 Higher dimensions

§7.3 Partitions of unity

§7.4 Approximative identities on $\mathbb{R}$

Now we will try to develop a more general way of approximating functions.

§7.5 Banach algebras

§7.6 Literature used for this chapter

Chapter 8

Differential Geometry, Lecture 1

Lemma 8.1.

Lemma 8.2. *The distance between two points on the evolvent is given by*

$$(c(t_1), c(t_2)) = \frac{|\kappa_\gamma(t_2) - \kappa_\gamma(t_1)|}{|\kappa_\gamma(t_1)\kappa_\gamma(t_2)|} = |r_1$$

Observation. Curvature is the determinant of the Frenet frame.

Observation. Walking the curve backwards negates the curvatues, but does not change its derivative!

Chapter 9

Differential Geometry, Lecture 2

§9.1 Curves in $\mathbb{R}^3$

Let $\gamma : (a, b) \to \mathbb{R}^3$ be parametrised by arclength and $\gamma'' \neq 0$. Then

$$\gamma' \perp \gamma''$$

and we can define

$$N := \frac{\gamma''}{||\gamma''||}.$$

We need a third vector, ideally so that the three vectors are positively oriented. No worries, we can put down

$$B := \tau \times N.$$

Definition 9.1. *We define the curvature in* $\mathbb{R}^3$ *to be*

$$\kappa_\gamma := ||\gamma''||$$

.

Definition 9.2. *We define* trójnóg Freneta *to be the ordered orthonormal basis*

$$(\tau, N, B).$$

We have an analogue of the Frenet equations

Theorem 9.1 (3D Frenet Equations)**.** *For the* Frenet trójnóg *we have*

$$\frac{\mathrm{d}}{\mathrm{d}t}(T, N, B) = (T, N, B)M$$

Proof. The first row follows from the definition. Let

$$A = (T, N, B).$$

This is an orthonormal matrix for all $t \in (a, b)$, so it traces a curve in SO(3) and we want its derivative. We will show

$$A^{-1}A'$$

is skew-symmetric – this will give us the theorem. We know that

$$AA^T = \mathrm{Id},$$

so let us differentiate both sides and obtain

$$(A')^T A + A^T A' = 0$$

and

$$(A^T A')^T = (A')^T A = -A^T A'.$$

Where we have used the matrix Leibniz rule. □

Definition 9.3. *The function τ mentioned in the previous theorem is called **torsion**.*

We have another analogue to curves in $\mathbb{R}^2$.

Theorem 9.2 (Fundamental Theorem of Curves is $\mathbb{R}^3$)**.** *For a positive function*

$$\kappa : (a, b) \to \mathbb{R}_+,$$

an arbitrary function

$$\tau : (a, b) \to \mathbb{R}$$

and any positively oriented orthonormal basis of $\mathbb{R}^3$ (the initial Frenet frame), there exists exactly one arclength parametrised curve

$$\gamma : (a, b) \to \mathbb{R}^3$$

which has Frenet frame at $t = 0$ equal to B.

Proof (sketch). As in the 2-dimensional case – put down a differential equation. □

What is torsion? The torsion describes how much a curve escapes a plane in which it starts (the *osculating plane*).

§9.2 Curvature of surfaces

Consider a surface

$$S \subseteq \mathbb{R}^3.$$

We will want to define its curvature in terms of plane curvature. We will work with surfaces given as immersions

$$\Sigma : U \to \mathbb{R}^3,$$

where $U \subseteq \mathbb{R}^2$.

Definition 9.4. *Let $p \in U \subseteq \mathbb{R}^2$ and Σ be as above. The **Riemann metric** on TU is the smooth, positive definite bilinear form given at p by*

$$\langle v_p, w_p \rangle_\Sigma = \langle \mathrm{d}\Sigma(v_p), \mathrm{d}\Sigma(w_p) \rangle .$$

Notation. We will denote by

$$\partial_x = \mathrm{d}\Sigma(\partial_x)$$
$$\partial_y = \mathrm{d}\Sigma(\partial_y)$$

the push by Σ of the standard vector field on $\mathbb{R}^2$.

Definition 9.5. *The **surface normal vector** to the surface Σ at p is*

$$n_p = \frac{\partial_x \times \partial_y}{||\partial_x \times \partial_y||}.$$

Now we can start talking about curvature!

Definition 9.6. *For $v \in T_p\Sigma$ of length 1, the **curvature***

$$\kappa_p(v)$$

is equal to the plane curvature κ at p of the curve, which is the intersection of Σ and the plane

$$\mathrm{Lin}\{n_p, v\}.$$

(This is a set – the parametrization can be recovered with aid of the inverse function theorem).

Since we have defined the above notion of curvature for $||v|| = 1$, we have a function

$$\kappa : S^1 \to \mathbb{R}$$

9.2.1 Example of surface curvature

Let us isometrically change the coordinate system so that $p = 0$ and take

$$T_p\Sigma = \mathrm{Lin}\{e_1, e_2\}.$$

Problem. This can be done so that in a neighbourhood of 0 the surface looks like the graph of

$$f = \frac{1}{2}\alpha x^2 + \frac{1}{2}\beta y^2 + O(x^3 + y^3).$$

Let $v = (\cos\theta, \sin\theta, 0)$. We have that

$$n_p = (0, 0, 1)$$

and so

$$\gamma_v(t) = f(tv) = \frac{1}{2}\alpha\cos^2\theta + \frac{1}{2}\beta\sin^2\theta + O(t^3) = ct^2 + O(t^3).$$

The curvature of this after going back to the plane is the same as the curve

$$g(t) = ct^2.$$

What is the curvature of $g(t)$ at $t = 0$? After applying a homothety of scale c^{-1} we get a curve $h(t) = t^2$, which has curvature 2 at $t = 0$. By considering osculating circles, the curvature of $g(t)$ is

$$2c = \alpha\cos^2\theta + \beta\sin^2\theta.$$

The extreme values of this are α, β obtained for $v = (1, 0, 0)$ and $v = (0, 1, 0)$. Note that the vectors are orthogonal!

9.2.2 Why will we multiply these?

Picture one – immersing a square into $\mathbb{R}^3$ as a cylinder.

§9.3 Euclidean connection

Take two vector fields X, Y and a surface Σ. We have

$$D_X Y = (X(Y_1), X(Y_2), X(Y_3)).$$

A part of that will be tangent to Σ, and part will be normal. Writing this decomposition down defines

$$D_X Y = \nabla_X Y + \mathrm{I\!I}(X, Y) n_p.$$

Lemma 9.1. *The second derivative of*

$$g \circ f$$

is given by

$$(g'' \circ f) \cdot (f')^2 + (g' \circ f) \cdot f''$$

Proof. By the Chain Rule,

$$(g \circ f)' = f' \cdot (g' \circ f).$$

To take the derivative of this, we use the product rule and chain rule. □

Chapter 10

Differential Geometry, Problemset 1

Problem 1

Solved in the lecture.

Problem 2

Solved in the lecture.

Problem 3

By 9.1, if the hypothesis holds for a curve γ, it also holds for all reparametrizations. Therefore, we may assume that γ is parametrized by arclength. Denote the center of the circle as c. We now have

$$\begin{aligned} f(\gamma) &= \langle \gamma - c, \gamma - c \rangle - R^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma) &= 2\langle \dot{\gamma}, \gamma - c \rangle \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\gamma) &= 2\langle \ddot{\gamma}, \gamma - c \rangle + 2\langle \dot{\gamma}, \dot{\gamma} \rangle \end{aligned}$$

Now, $f(\gamma) = 0$ iff $\gamma(t) \in C(x, R)$. Then, the first derivative is 0 iff $\dot{\gamma}$ is perpendicular to the radius of the circle. Note that the tangent to the circle has precisely this property, so this gives us equivalence of γ being tangent to the circle and the derivative of $f(\gamma)$ being $= 0$.

For the third part, we assume arclength parametrization. Then the equation becomes

$$0 = 1 + \langle \ddot{\gamma}, \gamma - c \rangle,$$

which means that the length of $\ddot{\gamma}$ must be $\frac{1}{R}$ and its direction must be opposite to $\gamma - c$, i.e. it must be directed toward the centre of the circle. This is true for the second derivative of the arclength parametrised circle, so the second derivative of $f(\gamma)$ is zero iff the second derivatives of γ and the arclength parametrization of the circle agree.

Problem 4

An equivalent definition:

Definition 10.1. *For an arclength-parametrized curve γ, the circle of best fit to γ at s_0 is the unique tangent circle with the same signed curvature at the point of contact.*

Since any tangent circle has the same tangent, is also has the same positive normal ($N(s)$), and the second derivative is the curvature times the normal by definition of curvature **??**.

Problem 6

Assume the curve is parametrized by arclength. We will compute how the distance from the center of the circle changes along the curve. Let the circle have curvature κ. Then the radius of the circle is $1/\kappa$ and its center is the point

$$\gamma(0) + \frac{1}{\kappa}N(0).$$

Then, the vector from the centre to a point on the curve is

$$r(s) := \gamma(s) - \gamma(0) - \frac{1}{\kappa}N(0).$$

This is clearly not changed by translating the whole configuration, so without loss of generality $\gamma(0) = 0$. In what follows, we use the dot product differentiation formula 8.1.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}||r(s)||^2 &= \frac{\mathrm{d}}{\mathrm{d}s}\langle r(s), r(s)\rangle \\ &= 2\langle \dot{r}(s), r(s)\rangle \\ &= 2\langle \gamma(s), \dot{\gamma}(s))\rangle - 2\left\langle \frac{1}{\kappa}N(0), \dot{\gamma}(s)\right\rangle.\end{aligned}$$

At $s = 0$ both terms come out to 0, so we need to compute another derivative to see what is going on. We have

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}s^2}||r(s)||^2 &= 4\langle \dot{\gamma}(s), \dot{\gamma}(s)\rangle + 4\langle \gamma(s), \ddot{\gamma}(s)\rangle - 4\left\langle \frac{1}{\kappa}N(0), \ddot{\gamma}(s)\right\rangle \\ &= 4 + 4\langle \gamma(s), \ddot{\gamma}(s)\rangle - 4\frac{\kappa_\gamma(s)}{\kappa}\langle N(0), N(s)\rangle,\end{aligned}$$

where we have used **??** and **??**. At $s = 0$ this comes out to

$$4\left(1 - \frac{\kappa_\gamma(0)}{\kappa}\right),$$

which is positive for $\kappa_\gamma(0) < \kappa$ and negative for $\kappa_\gamma(0) > \kappa$. This concludes the problem: for example, if the curve γ stays inside the circle (even locally!), then we cannot have $\kappa_\gamma(0) < \kappa$, because then the distance would be increasing (by Taylor **??**).

Remark. This still works for negative κ. The only thing we need to check to make sure of that is that the center of the circle is where it is. It gets more tricky for $\kappa = 0$ – in that case the appropriate reformulation of *inside* and *outside* is on one or the other side of the line, and instead of the distance we should consider

$$\langle \gamma(s) - \gamma(0), N(0) \rangle.$$

The derivative of this is

$$2\langle \dot\gamma(s), N(0) \rangle,$$

which equals 0 at $s = 0$. The second derivative is

$$4\langle \ddot\gamma(s), N(0) \rangle = 4\kappa_\gamma(s)\langle N(s), N(0) \rangle$$

by Frenet **??**. At s this is just $4\kappa_\gamma(0)$, and analysing signs as above finishes the problem.

Problem 5

Recall from Problem 6 that

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}||r(s)||^2 = 4 + 4\langle \gamma(s), \ddot\gamma(s) \rangle - 4\frac{\kappa_\gamma(s)}{\kappa}\langle N(0), N(s) \rangle$$

and that the first derivative disappears at $s = 0$ for a tangent circle. For a circle of best fit, the curvature $\kappa = \kappa_\gamma(0)$ (see my solution of Problem 4), so the second derivative disappears as well. Computing the derivative of this

$$\frac{\mathrm{d}^3}{\mathrm{d}s^3}||r(s)||^2 = 4\langle \gamma(s), \dddot\gamma(s) \rangle + 4\langle \dot\gamma(s), \ddot\gamma(s) \rangle - 4\frac{\kappa'_\gamma(s)}{\kappa}\langle N(0), N(s) \rangle - 4\frac{\kappa_\gamma(s)}{\kappa}\langle N(0), \dot N(s) \rangle$$

For $s = 0$ this gives

$$\frac{\mathrm{d}^3}{\mathrm{d}s^3}||r(s)||^2 = -4\frac{\kappa'_\gamma(s)}{\kappa}.$$

So, if the derivative is nonzero

Problem 7

I have a shitty calculatory solution, but nothing nice geometrically. It goes like

1. Parametrize the points on the parabola by (p, p^2).
2. Centers of tangent circles lie on $p - 2pt, p^2 + pt$ and have radii $t\sqrt{1 + 4p^2}$.
3. Together with $y = x^2$ this gives an equation for the intersection points. Divided by $(x - p)^2$ the equation is
$$1 + (x + p)^2 + 2t = 0.$$
4. To avoid having two solutions, pick t so that $x = p$ is a solution of this. Then, $x = -3p$ is another solution (same value of the square).
5. Caclulate that in this configuration
$$t = -\frac{1 + 4p^2}{2}.$$

6. The tangent to the parabola at the chosen point is $(1, -6p)$. The point is $(-3p, 9p^2)$ and the center of the circle is $(2p + 4p^3, -p^2 - 1/2)$. The vector from the circle to the point is then $(5p + 4p^3, -10p^2 - 1/2)$.

7. The dot product between the tangent and the vector from the point to the center is a nonzero polynomial, so for almost all p the circle is not tangent.

Problem 8

Geometrically, the curve rolls into itself, like a spiral. Note that the curvature can also be negative or zero at some points, but increasing. In what follows, we consider a point of self-intersection.

10.0.1 Non-zero curvature

By the **??**, we know that if the curvature at some point is nonzero, then points in the direction of the absolute value of curvature are properly inside the osculating circle. Therefore, there can be no self-intersection at a point of nonzero curvature.

10.0.2 Zero curvature

For a point of zero curvature, we know from the remark to problem 6 that the first derivative of the signed distance from the line (which is the osculating circle at a point of curvature 0) is 0 and the second derivative is strictly positive after the point of zero curvature and strictly negative before (in a small neighbourhood!).

Taking points arbitrarily close to zero, we see that

TO BE CONTINUED!

Problem 9

Chapter 11

Measure Theory Bank of Lemmas

Lemma 11.1 (Generating a *sigma*-algebra)**.** *Fix a space X. For any family of sets $\mathcal{A}$, $\sigma(\mathcal{A})$ can be generated by any of the following sets of operations:*

1. *complements and countable unions.*

Lemma 11.2. *Let $\mathcal{B}$ be a basis for the topology of X. Then*

$$\sigma\left(\mathcal{B}\right) = \operatorname{Bor} X.$$

Lemma 11.3. *The set operations and the measure taking operations are continuous with respect to the symmetric difference pseudometric.*

Lemma 11.4. *A bounded measurable function f on a measure space (X, μ) can be uniformly approximated by simple functions.*

Proof. Let

$$|f| \leqslant [-M, M].$$

We will construct the approximation by considering *bins* of the values of f, i.e. the sets

$$A_k := f^{-1}\Big[[k\varepsilon, (k+1)\varepsilon)\Big]$$

for $k \in \mathbb{Z}$. In such a *bin*, all the values are within an ε of each other. Since f is bounded, all but finitely many of the bins are empty X, so the function

$$\widetilde{f}_\varepsilon := \sum_{A_k \neq \varnothing} (k\varepsilon) \cdot \chi_{A_k}$$

□

Remark. This works equally well for almost everywhere bounded functions, giving almost everywhere uniform convergence.

Lemma 11.5. *Let A, A_1, ..., A_k be measurable sets such that*

$$\forall k.\, \mu(A_i \cap A) \geq (1 - \delta_i)\mu(A).$$

Then

$$\mu(A \cap A_1 \cap A_2 \cap \ldots \cap A_k) \geq \Big(1 - \sum \delta_i\Big)\, \mu(A).$$

Proof. Union bound on the sets
$$A \cap A_i^c.$$
□

Remark. This also works for an infinite sequence of sets A_k; we obtain.
$$\mu\left(A \cap \bigcap_k A_k\right) \geqslant \left(1 - \sum_k \delta_k\right) \mu(A).$$

Lemma 11.6. *Let (X,d) be a metric space, $x \in X$ and $A \subseteq X$ be closed. Then there exists an $a \in A$ such that*
$$d(x,a) = d(x,A).$$

Chapter 12

An introduction to geometric measure theory

In this chapter, we study the links between the topology and geometry of $\mathbb{R}$ and the Lebesgue measure. We first give two examples of how the two structures agree, and one example of how they don't.

Isometries. Consider the group $\operatorname{Isom}\mathbb{R}$ of the isometries of $\mathbb{R}$ with the euclidean metric. One easily shows that this group consists of functions of the form

$$x + a \text{ or } a - x$$

for $a \in \mathbb{R}$. The Lebesgue measure is invariant on transformations $g \in \operatorname{Isom}\mathbb{R}$, i.e.

$$\lambda(gA) = \lambda(A)$$

for all measurable $A \subseteq \mathbb{R}$. A corollary of this is that the Lebesgue measure is invariant w.r.t. the addition operation on $\mathbb{R}$, which gives the reals the structure of a topological group.

Affine transformations. Similarly to the above, the Lebesque measure work well with the action of the affine transformation group $\operatorname{Aff}\mathbb{R}$. Directly from the definition, the group of affine transformations consists of the functions

$$g_{a,b}(x) := ax + b$$

for $r \neq 0$, and the interaction with measure is given by

$$\lambda(g_{a,b}A) = |a| \cdot \lambda(A).$$

Topology. There is a disconnect between the topological (nonempty interior) and measure-theoretic (positive measure) notions of *large* or *non-negligable* – the topological notion is strictly stronger! Indeed, a set with nonempty interior has positive measure, but if we enumerate the rationals as

$$\mathbb{Q} = \{q_1, q_2, \ldots\}$$

the set

$$\mathbb{R} \setminus \bigcup_{n=1}^{\infty} (q_n - \frac{\varepsilon}{2^{n+1}}, q_n + \frac{\varepsilon}{2^{n+1}})$$

has comeasure ε, but is nowhere dense.

However, there does exist a link between the two notions. It is a bit more subtle.

Definition 12.1. *Fix a measurable set $A \subseteq \mathbb{R}$. A point $x \in \mathbb{R}$ is called a* ***density point*** *iff*

$$\lim_{\delta \to 0^+} \frac{\lambda(A \cap B(x, \delta))}{2\delta} = 1.$$

The 2δ in the numerator is of course $\lambda(B(x, \delta))$.

Definition 12.2. *The set of density points of A will be denoted $\phi(A)$.*

Note that a density point is by neccesity an accumulation point. The promised link between geometry, measure and topology is provided by the theorem below.

Ankified **Theorem 12.1** (Lebesgue Density Theorem)**.** *Let $A \subseteq \mathbb{R}$ be a measurable set. Then almost all points of A are density points of A in the sense that*

$$\lambda^*(A \setminus \phi(A)) = 0.$$

Remark. Note that the theorem follows trivially for null sets. Also, for a given A, we may as well apply the theorem to A^c to get that almost all points outside of A have density 0.

For the proof of the **Lebesgue Density Theorem**, we will need a tool, which we introduce now and prove later.

Definition 12.3. *A family $\mathcal{J}$ of nontrivial closed intervals is called a* ***Vitali cover*** *of a set A (not necessarily measurable) if for any given $\varepsilon > 0$ and $x \in A$ there is an interval $J \in \mathcal{J}$ such that*

$$\operatorname{diam} J < \varepsilon \wedge x \in J.$$

In particular

$$A \subseteq \bigcup \mathcal{J}.$$

Ankified **Theorem 12.2** (Vitali Covering Theorem)**.** *If $\mathcal{J}$ is a Vitali cover of A, there exists a sequence of pariwaise disjoint segments $J_n \in \mathcal{J}$ such that*

$$\lambda\left(A \setminus \bigcup_n J_n\right) = 0.$$

Why is this theorem useful? Vitali's theorem may not sound very smart on first glance. Its strength lies in the *disjointness* of the cover. If we go about choosing the cover J_n without any guarantees, we can for example choose

$$\bigcup_{q_n \in \mathbb{Q}} \left(q_n - \frac{\varepsilon}{2^{n+1}}, q_n + \frac{\varepsilon}{2^{n+1}}\right)$$

and get stuck! We have only covered a subset of the reals of size ε, but we cannot use any other segment by density of $\mathbb{Q}$.

Proof of the Lebesgue Density Theorem. We represent

$$A \setminus \phi(A) = \bigcup_k A_k$$

for

$$A_k = \left\{ a \in A : \liminf_{\delta \to 0^+} \frac{\lambda(A \cap B(a,\delta))}{2\delta} < 1 - \frac{1}{k} \right\}.$$

It suffices to show

$$\lambda^*(A_k) = 0$$

for all k to finish the proof. Since we may represent A as

$$A = \bigcup_{z \in \mathbb{Z}} A \cap [z-1, z+1]$$

and being a density point of A is the same as being a density point of one of the *cutouts* in the sum above, we may assume without loss of generality that $A \subseteq [0,1]$.

By definition of outer measure, we can approximate A_k from above by an open set U such that

$$\lambda^*(A_k) \leqslant \lambda(U) \leqslant \lambda^*(A_k) + \varepsilon.$$

Construct a covering

$$\mathcal{J} = \left\{ [a,b] : [a,b] \subseteq U,\ \lambda\Big(A \cap [a,b]\Big) \leqslant \left(1 - \frac{1}{k}\right) \lambda[a,b] \right\}.$$

It is a Vitali cover of A_k. By Vitali's Theorem we can pick a pairwise disjoint sequence of intervals $J_i \in \mathcal{J}$ for which

$$\lambda^* \left(A_k \setminus \bigcup_i J_i \right) = 0.$$

This gives

$$\begin{aligned}
\lambda^*(A_k) &= \lambda^* \left(A_k \cap \bigcup_i J_i \right) \\
&\leqslant \sum_i \lambda^*(A_k \cap J_i) \\
&\leqslant \sum_i \lambda^*(A \cap J_i) \\
&\leqslant \left(1 - \frac{1}{k}\right) \sum_i \lambda(J_i) \\
&\leqslant \left(1 - \frac{1}{k}\right) \lambda(U) \\
&\leqslant \left(1 - \frac{1}{k}\right) \left(\lambda^*(A_k) + \varepsilon\right).
\end{aligned}$$

The passage from line 2 to 3 may seem trivial, but is in fact crucial. This is the place where we use $A_k \subseteq A$! Otherwise the theorem is quite absurd, even for simple examples like $[0,1]$. Since $\lambda^*(A_k) \leqslant \lambda(A) < \infty$, we can rearrange this to obtain

$$\lambda^*(A_k) \leqslant (k-1)\varepsilon.$$

Since ε can be picked arbitrarily close to 0, we get

$$\lambda^*(A_k) = 0.$$

$\square$

Proof of the Vitali Covering Theorem. The key to avoiding the *trap* we wrote about after stating the VCT is to choose the segments to be as large as possible – or at least not embarassingly small.

Without loss of generality, A is bounded since we can sum the coverings of $A \cap (n, n+1)$. The sequence of segments we pick is denoted J_n. In that case we may also assume $\bigcup \mathcal{J}$ is bounded. Its prefixes are

$$\begin{aligned} P_n &:= \bigcup_{i<n} J_i \\ \mathcal{J}_n &:= \{J \in \mathcal{J} : J \cap P_n = \varnothing\} \end{aligned}$$

and the *width* of what we can choose is

$$\gamma_n := \sup_{J \in \mathcal{J}_n} \operatorname{diam} J.$$

Note that in particular

$$\begin{aligned} P_1 &= \varnothing, \\ \mathcal{J}_1 &= \mathcal{J} \\ \gamma_1 &\leqslant \operatorname{diam} A < \infty. \end{aligned}$$

At each step, we choose J_n so that

$$\operatorname{diam} J_n \geqslant \frac{1}{2}\gamma_n,$$

or we stop if $\gamma_n = 0$ at some point.

Claim 1. The sequence γ_n converges monotonically to 0.

Proof of Claim 1. Being the supremum of ever decreasing sets, γ_n is decreasing. It is also non-negative, so the sequence converges and $\lim_n \gamma_n \geqslant 0$. Suppose that $\lim_n \gamma_n = c > 0$. Then in the construction, we would almost always choose disjoint intervals of diameter at least $c/2$. This is impossible, since $\bigcup \mathcal{J}$ was assumed to be bounded, so it has finite measure! $\square$

The key to proving that the choice procedure is correct will be the **blowup**, which we define for $J = [x-r, x+r]$ as

$$\widetilde{J} := [x - 5r,\, x + 5r]$$

Claim 2. At all steps of the construction

$$A \subseteq \mathcal{J}_n \cup \bigcup_{i \geqslant n} \widetilde{J}_i.$$

Proof of Claim 2. The set $\mathcal{J}_n$ is closed as a union of closed intervals. Therefore, if $a \in A \setminus \mathcal{J}_n$, there is a nondegenerate interval $I \ni a$. Since $\gamma_n \to 0$ by **Claim 1**, I is not considered in the construction of the sequence J_n for almost all n. Let n_0 be the last step where it is considered. Then we must have $I \cap J_{n_0+1} \neq 0$, because that is the step at which I is no longer considered.

We will show that this implies $a \in \widetilde{J}_{n_0+1}$. Let $J_{n_0+1} = [x - r, x + r]$ and $y \in I \cap J_{n_0+1}$. Then we have

$$\begin{aligned} d(a, x) &\leqslant d(a, y) + d(y, z) \\ &\leqslant \operatorname{diam} I + r \\ &\leqslant (2 \cdot \operatorname{diam} J_{n_0+1}) + r \\ &= 2 \cdot 2r + r \\ &= 5r, \end{aligned}$$

where the diameter bound comes from the definition of γ_{n_0} and the fact that I is still available at step n_0 of the construction. $\square$

To finish the proof of Vitali's Covering Theorem, we compute that for all n

$$\begin{aligned} \lambda^* (A \setminus P_n) &\leqslant \lambda^* \left(A \cap \bigcup_{i \geqslant n} \widetilde{J}_i \right) \\ &\leqslant \lambda \left(\bigcup_{i \geqslant n} \widetilde{J}_i \right) \\ &\leqslant \sum_{i \geqslant n} \lambda(\widetilde{J}_i) \\ &= 5 \sum_{i \geqslant n} \lambda(J_i). \end{aligned}$$

These are the tails of the convergent series

$$\sum_{i=1}^{\infty} \lambda(J_i) = \lambda \left(\bigcup_i J_i \right) < \infty,$$

so we get

$$\lambda^* \left(A \setminus \bigcup_i J_i \right) \leqslant \lambda^* (A \setminus P_n) \to 0.$$

$\square$

Remark. Retracing the argument behind **Claim 2.**, we might prove that for any $\alpha < 1$, if we define γ_n with a coefficient of α instead of $\frac{1}{2}$, the constant used for blowing up intervals can be brought down to

$$1 + \frac{2}{\alpha}.$$

In particular, we can get arbitrarily close to 3.

§12.1 Corollaries and the Lebesgue Differentiation Theorem

Theorem 12.3 (Lebesgue Differentiation Theorem)**.** *Let $f \in L^1(\mathbb{R})$. Then, for almost all x,*

$$\lim_{\delta\to 0^+} \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} f(s)\,\mathrm{d}\lambda(s) = f(x).$$

Proof. For characteristic functions, this is just a restatement of the Lebesgue Density Theorem. $\square$

§12.2 Generalization to metric spaces

The argument in the proof of the VCT was written so that it is easliy generalizable to any metric space with a measure on its Borel sets.

To be more precise, what we need to lift the argument is that

$$\mu\left(B(x,5r)\right) \leqslant C\mu\left(B(x,r)\right)$$

for some constant C. We can also substitute any constant larger that 3 instead of 5.

Chapter 13

One (Cantor) set to rule them all

§13.1 Ternary Cantor

Let us begin by making a construction. Take the closed interval $C_0 := [0, 1]$ and remove the middle one third of it in such a way that the remaining two interval are closed. The result of this is

$$C_1 := \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Now, repeat the operation of cutting out the middle third and call the result C_2. We can repeat this *ad inifinitum* and obtain a decresing sequence of sets

$$[0, 1] = C_0 \supset C_1 \supset C_2 \supset \dots.$$

Perhaps surprisingly, there are numbers which are not removed at any step, i.e. the intersection

$$\mathcal{C}_3 := \bigcap_{k=0}^{\infty} C_k$$

is nonempty! It contains 0 and 1. In fact, any number which can be written in base 3 using only 0's and 2's is an element of this intersection. These are in fact all such numbers. We introduce a tool to prove that.

Lemma 13.1. *Let $b \geqslant 2$ be a positional system base and x_0 be a number with k digits after the positional point. Then, the numbers formed by adjoining (perhaps infinitely many) digits to the base b representation of x_0 are all the numbers in the interval*

$$\left[x_0,\, x_0 + b^{-k}\right].$$

If we allow only finite extensions, we get the b-ary rational numbers in that interval, and if we disallow the infinite extension by the digit $(b-1)$, we get the interval

$$\left[x_0,\, x_0 + b^{-k}\right).$$

Lemma 13.2. *A real number $x \in [0, 1]$ is an element of $\mathcal{C}_3$ iff x can be written in base using only the digits 0 and 2.*

Proof (if). We proceed by induction with the induction thesis: *x belongs to $\mathcal{C}_3$ iff x can be written in base 3 so that its first k digits are 0 or 2.* It should be clear that this thesis is equivalent to the lemma statement. For each k, the statement is true by the previous lemma. □

§13.2 Abstract Cantor

The ternary Cantor set has many interesting properties. However, to study it, we will move to a more convenient representation. We can think of the Cantor set as the set of leaves of an infinite binary tree – starting from the root, at each level we choose whether to go right or left, or whether to insert 0 or 2 as the next digit in the base-3 representation of an $x \in \mathcal{C}_3$.

In this way, we can represent the ternary Cantor as

$$\mathcal{C} := \{0,1\}^{\mathbb{N}}.$$

That the map we just described is a bijection follows from

Lemma 13.3. *Let $d_k, \widetilde{d_k}$ be two sequences of base b digits. Then the corresponding real numbers are equal iff d_k and $\widetilde{d}$ agree on some prefix and afterwards one of them is 0 and the other is $(b-1)$.*

Proof. The condition implies equality of numbers by the sum of a geometric series. The other direction follows by looking at the first moment the expansions differ at then bounding the series sum. □

What is missing from this description is the topology. We topologize $\mathcal{C}$ by the metric

$$d(x,y) = \begin{cases} 0 & \text{for } x = y \\ \frac{1}{n} & \text{for } x \neq y, \end{cases}$$

which can also be written succintly as

$$d(x,y) = \frac{1}{n_0(x,y)}$$

with the notation

$$n_0(x,y) := \inf\{n : x_n \neq y_n\}$$

for the first index at which x and y differ. The function d may not look like a metric at first sight, but in fact it has an even better property.

Lemma 13.4. *For the metric d described above we have for all $x, y, z \in \mathcal{C}$*

$$d(x,z) \leqslant \max\{d(x,y), d(y,z)\}.$$

*In particular, d is an **ultrametric**.*

Proof. Recall that $n_0(x,z)$ is the first position at which x and z differ. Then any y has to differ with at least one of y and z at n_0, but might even earlier. This gives

$$n_0(x,z) \geqslant \min\big(n_0(x,y), n_0(y,z)\big).$$

Since the function $x \mapsto 1/x$ is decreasing, the thesis follows. □

We have established that $(\mathcal{C}, d)$ is a metric space. It is, in fact, homeomorphic with the subspace topology of $\mathcal{C}_3$ inherited from $[0,1]$.

Lemma 13.5. *The function*

$$h_3 : \mathcal{C} \to \mathcal{C}_3$$

defined by

$$h_3(x) := \sum_{k=1}^{\infty} \frac{2x_k}{3^k}$$

is a homeomorphism.

Proof. Bijectivity follows from the number-system lemma 13.3 and 13.2. For continuity, put down $n_0 := n_0(x, y)$ and compute

$$\begin{aligned} |h_3(x) - h_3(y)| &= \sum_{k=1}^{\infty} \frac{2|x_k - y_k|}{3^k} \\ &\leqslant \sum_{k=n_0}^{\infty} \frac{2}{3^k} \\ &= \frac{2}{3^{n_0}} \cdot \frac{3}{2} \\ &= \frac{1}{3^{n_0-1}}. \end{aligned}$$

The continuity of the inverse follows from the bound

$$|h_3(x) - h_3(y)| \geqslant \frac{2}{3^{n_0}}.$$

□

The function h_3 in 13.5 can be understood as a base 3 expansion operator. When we consider a base 2 expansion instead, we lose bijectivity, but we can cover the whole interval.

Lemma 13.6. *The function*

$$h_2 : \mathcal{C} \to [0, 1]$$

given by

$$h_2(x) := \sum_{k=1}^{\infty} \frac{x_k}{2^k}$$

is a continuous surjection.

Proof. Surjectivity follows from number system properties, and continuity is essentialy the same calculation as in the proof of 13.5. □

Theorem 13.1 (The Universal Property of the Cantor Set)**.** *Every metrizable compact space K is a continuous image of $\mathcal{C}$.*

Proof. Considering an element of $\mathcal{C}$ as a binary expansion, we have by 13.6 a surjection

$$h_2 : \{0, 1\}^{\mathbb{N}} \twoheadrightarrow [0, 1].$$

The space K can be embedded into the Hilbert cube by the **Urysohn Metrization Theorem ??**. By compactness of K, the image of the embedding is a compact and thus a closed subset. We also have a surjection

$$h : \{0, 1\}^{\mathbb{N}} \to [0, 1]^{\mathbb{N}}$$

by using the previous surjection and *unweaving* the Cantor set into the product of countably many Cantor sets, i.e. using

$$\mathbb{N} \cong \mathbb{N} \times \mathbb{N} \implies \mathcal{C} = \{0,1\}^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}\times\mathbb{N}} \cong \left(\{0,1\}^{\mathbb{N}}\right)^{\mathbb{N}} = \mathcal{C}^{\mathbb{N}}.$$

The last step is using the fact that any closed set of $\mathcal{C}$ is a retract of $\mathcal{C}$, which is **??**. □

A warning against generalization. If K is a compact set, it embeds into a *Tichonov Cube*

$$K \to [0,1]^{\Gamma}$$

and we can surject the Tichonov cube with a generalized Cantor set

$$\{0,1\}^{\Gamma},$$

but the universality theorem fails!

§13.3 Topology of the Cantor set

Definition 13.1 (Cantor Cylinder)**.** *Let*

$$\varphi : \mathbb{N} \multimap \{0,1\}$$

be a partial function with finite domain. Then we define the ***cylinder set*** *with base φ as*

$$[\varphi] := \{x \in \{0,1\} : x|_I = \varphi\}.$$

Lemma 13.7. *The sets $[\varphi]$ form a base of the topology of $\{0,1\}^{\mathbb{N}}$.*

Definition 13.2. *A set $A \subseteq \mathcal{C}$ is* ***determined*** *by $I \subseteq \mathbb{N}$, which we donte by $A \sim I$ if for all $x \in A$, $y \in \mathcal{C}$ we have*

$$x|_I = y|_I \implies y \in A.$$

Equivalently,

$$\pi_I^{-1}\pi_I[A] = A.$$

Lemma 13.8 (Clopen sets in the Cantor set)**.** *A set $A \subseteq \mathcal{C}$ is clopen iff $A \sim I$ for some finite $I \subseteq \mathbb{N}$. In particular, clopen sets can be written as a finite union of disjoint basis clopens $[\varphi_i]$ for φ_i with finite domain.*

(Direction one). If A is clopen, then

$$A = \bigcup_i [\varphi_i]$$

for some finitely many (by compactness) φ_i with finite domain I_i. Then

$$A \sim \bigcup_i I_i.$$

□

(Other direction). if $A \sim I$, blabla □

Immediately, a lemma follows.

Lemma 13.9 (Cantor set is zerodimensional)**.** *The Cantor set $\mathcal{C}$ is zerodimensional, i.e. it has a base of clopen sets.*

Theorem 13.2 (Topological characterisation of the Cantor set)**.** *If a topological space K is compact, metrizable, zerodimensional with no isolated points, then*

$$K \cong C.$$

§13.4 The group structure

The Cantor set has a natural abelian group structure given by its product structure. We can phrase it even more efficiently when we think of $\mathcal{C}$ as $\mathcal{P}(\mathbb{N})$ – the symmetric difference (or xor for the informatically inclined).

$$A \oplus B := A \Delta B$$

Every element has order two!

Fact. Together with the operation $\oplus$, the Cantor set $\mathcal{C}$ is a compact topological group, i.e. the function

$$(x, y) \mapsto x \oplus y$$

is continuous (in general the second element is inversed, but here every element is its own inverse anyway).

§13.5 Measure

We can define the measure on the Cantor set as a countable product of probability measures:

$$\nu = \bigotimes_{n=1}^{\infty} (\frac{1}{2}(\delta_0 + \delta_1)).$$

But we will do it by hand.

Definition 13.3. *Let $A \subseteq \mathcal{C}$ be clopen. Then*

$$A \sim \{1, 2, \ldots, n\}$$

for some n. Let

$$A' := \pi_{\{1,2,\ldots,n\}}[A]$$

We define its measure to be

$$\nu(A) := \frac{\#A'}{2^n}.$$

This makes sense with the probabilistic definition.

Theorem 13.3 (Well-definedness of the premeasure)**.** *The function*

$$\nu : \operatorname{Clop} \mathcal{C} \to \mathbb{R}$$

is a well-defined, additive function on the set algebra $\operatorname{Clop} \mathcal{C}$*.*

Proof. Since the Cantor set is compact, ν is automatically downward continuous on the empty set. By Caratheodory's Theorem, ν extends uniquely to a probabilistic measure on

$$\operatorname{Bor}\mathcal{C} = \sigma\left(\operatorname{Clop}\mathcal{C}\right).$$

What now??

$$\mathcal{A} := \left\{B \in \operatorname{Bor}\mathcal{C} : \forall \varepsilon > 0. \exists A \in \operatorname{Clop} C. \nu(A \Delta B) < \varepsilon\right\}.$$

We prove that this is a σ-algebra.

There is a nice formula for cylinders.

Lemma 13.10 (Measure of a cylinder)**.** *For a partial function*

$$\varphi : \mathbb{N} \multimap \{0, 1\},$$

its cylinder has measure

$$\nu\left[\phi\right] = 2^{-|\operatorname{dom}\varphi|}.$$

The result holds even if $\operatorname{dom}\varphi$ *is infinite, in which case the measure is* 0*.*

Proof. For finite-domain partial functions ϕ, take

$$\operatorname{dom}\varphi =: I \subseteq \{1, 2, \ldots, n\} =: [n]$$

for some n. Then

$$\left|\pi_{[n]}[\varphi]\right| = \frac{2^{n-|I|}}{2^n} = 2^{-|I|}.$$

For infinite-domain functions ϕ, taking a decreasing intersection

$$[\phi] = \bigcap_n \left[\phi|_{[n]}\right]$$

shows that the measure of the intersection is 0. □

Theorem 13.4. *The measure* ν *is the Haar measure on* $\mathcal{C}$*, that is, the unique probability measure invariant under group actions*

$$\nu\left(x \oplus B\right) = \nu(B)$$

for all $x \in \mathcal{C}$*,* $B \subseteq \mathcal{C}$*.*

Proof. Let us first consider $B = [\varphi]$, and $I = \operatorname{dom}\varphi$. Then

$$\nu\left(x \oplus [\varphi]\right) = \nu\left([x \oplus \varphi]\right) = \nu\left([\varphi]\right).$$

A clopen is a disjoint sum of $[\varphi_i]$ for finitely many φ_i, so additivity on clopens follows. Now, take a superficially different measure

$$\nu_x(B) := \nu\left(x \oplus B\right).$$

Since ν and ν_x agree on clopens, by uniqueness in Caratheodory's Theorem they agree on all sets. □

Note the isomorphism

$$(C, \oplus) \cong (\mathcal{P}(\mathbb{N}), \Delta)$$

of (topological) groups.

§13.6 Normal number theorem

Definition 13.4. *Let $A \subseteq \mathcal{C}$. We call A a* ***tail*** *set if*

$$A \sim \{k : k \geqslant n\}$$

for all n. Equivalently, if $x \in A$ and $x(n) = y(n)$ for almost all n, then $y \in A$.

Example. A naturally occuring example of a tail set is

$$A_\beta := \left\{ x \in \mathcal{C} : \lim_n \frac{x(1) + \ldots x(n)}{n} = \beta \right\}.$$

Theorem 13.5 (Kolmogorov zero-one law for the Cantor set)**.** *A borel tail set $A \subseteq \mathcal{C}$ has measure 0 or 1.*

Proof. Take a basis set $[\varphi]$. We have

$$\nu\left([\varphi] \cap A\right) = \nu\left([\varphi]\right) \cdot \nu(A).$$

From this immediately follows that this work for any $B \in \operatorname{Clop} \mathcal{C}$. Now approximate A by a clopen B so that

$$\nu\left(A \Delta \mathcal{C}\right).$$

To finish the proof, compute

$$\nu(A) \cdot \nu(B) = \nu(A \cap B) \geqslant \nu(A) - \varepsilon\nu(A).$$

□

Returning to the example we have $\nu(A_\beta) \in \{0, 1\}$. We have

$$\nu(A_\beta) = \nu(A_{1-\beta}).$$

Theorem 13.6 (Borel's normal number theorem)**.**

$$\nu\left(A_{\frac{1}{2}}\right) = 1.$$

Remark. According to Billingsley, this theorem was the founding work of modern probability theorem, which is founded on limit theorems.

Chapter 14

Measures on Topological Spaces, Problemset 1

Problem 8

Problem 8a

Problem 4

Extension 1

We show that the set can be the graph of a function! Let Z be a borel set of positive measure and define

$$T_Z = \{x : \lambda(Z_x) > 0\}.$$

Then T_Z is a measurable set by Fubini's Theorem. We can pick a compact subset T'_Z. A compact set of positive measure has at least $\mathfrak{c}$ elements, and there are as many borel sets. Then, enumerate borel sets of $\mathbb{R}^2$.

Chapter 15

Measures on Topological Spaces, problemset 2

Problem 1

Such an a exists by 11.6. If we have two a_1, a_2 such that

$$\rho(x, a_1) = \rho(x, a_2)$$

then a_1, a_2 must agree and disagree with x at all places, so in fact $a_1 = a_2$, thus r_A is well-defined. For any $a \in A$, $d(a, A) = 0 = d(a, a)$, so r_A is a retraction. What remains to be shown is continuity.

Let x, y agree up to $n_0(x, y)$. Then $r_A(x)$ and $r_A(y)$ also agree up to $n_0(x, y)$ – if they differed earlier, we could use $r_A(x)$ instead of $r_A(y)$ and get a closer point a in the definition! So we have

$$n_0(x, y) \leqslant n_0\left(r_A(x), r_A(y)\right)$$

and

$$d(x, y) \geqslant d\left(r_A(x), r_A(y)\right).$$

Remark. The metric $d(x, y) = 1/n_0(x, y)$, i.e. the first moment where x and y differ, won't work, because it can't tell apart points from which x differs at the same position!

Problem 2

Problem 3

Any $A, B \in \operatorname{Clop} \mathcal{C}$ can be written as disjoint sums of the basis sets $[\varphi]$ by 13.8. Since the condition distributes over disjoint sums, we will prove the statement for $A = [\varphi]$ and $B = [\psi]$ with

$$\left|\operatorname{dom} \varphi\right|, \left|\operatorname{dom} \psi\right| < \infty.$$

Let $I = \operatorname{dom} \varphi$, $J = \operatorname{dom} \psi$ be the disjoint(!) domains of φ, ψ. There is a function τ on $I \cup J$ such that

$$\tau|_I = \phi,\ \tau|_J = \psi.$$

For such a function,

$$[\varphi] \cap [\psi] = [\tau] \, .$$

Now take an n such that $I \cup J \subseteq \{1, 2, \ldots, n\}$ and denote the last set as $[n]$. By 13.10 we compute

$$\begin{aligned} \nu\,[\varphi] &= 2^{-|I|} \\ \nu\,[\psi] &= 2^{-|J|} \\ \nu\,[\tau] &= 2^{-|I \cup J|}, \end{aligned}$$

and $|I \cup J| = |I| \cup |J|$ finishes the proof. Now take arbitrary $A, B \in \operatorname{Bor} \mathcal{C}$ such that $A \sim I$, $B \sim J$. Approximate A, B by clopens A', B' to within an ε, i.e. so that

$$\nu(A \Delta A'),\, \nu(B \Delta B') < \varepsilon.$$

We cannot use the clopen statement we just proved since a priori A' and B' could be determined by sets with nonempty intersection. We can, however, improve the approximation with

$$\widetilde{A} := \pi_I^{-1} \pi_I A'.$$

The set $\widetilde{A}$ is still a clopen – since A' was determined by a finite set K, $\widetilde{A}$ is determined by $K \cap I$. Additionally we have

$$\widetilde{A} \Delta A \subseteq A' \Delta A,$$

so we have improved the approximation! Now, do the same for B' and use the statement for clopens to finish up the solution.

Warning! The reasoning below does not work! (For tail sets, for example)

We can approximate A, B by decreasing sequences of clopens by putting down

$$A_n := \pi_{[n]}^{-1} \pi_{[n]} A$$

and the same for B_n. We also approximate their intersection by decreasing clopens in the same way, i.e.

$$C_n := \pi_{[n]}^{-1} \pi_{[n]} (A \cap B).$$

For these approximations

$$C_n = A_n \cap B_n,$$

so by the first subproblem

$$\nu(C_n) = \nu(A_n \cap B_n) = \nu(A_n) \cdot \nu(B_n).$$

Since the measure ν is probabilistic, and hence continuous, by passing to the limit $n \to \infty$ we get what we need.

Problem 6

Any clopen $C \in \operatorname{Clop} \mathcal{C}$ is a disjoint sum of basis cylinders by 13.8. Since $\oplus$ is a group operation, the function

$$l_x(y) = x \oplus y$$

is bijective, so on the level of sets l_x distributes over disjoint sums. We check the property for a cylinder $[\varphi]$. This is easy, since

$$\nu\Big(x \oplus [\varphi]\Big) = \nu[x \oplus \varphi] = 2^{-|\operatorname{dom}\varphi|} = \nu[\varphi]$$

by 13.10. Now consider the family of sets

$$\mathcal{A} := \Big\{A : \forall x \in \mathcal{C}.\, \nu(A) = \nu(x \oplus A)\Big\}.$$

We will show that this is a σ-algebra. Since we have already shown that it contains all the clopens, which form a basis of the topology on $\mathcal{C}$, it will automatically be equal to $\operatorname{Bor}\mathcal{C}$ by 11.2.

A σ-algebra can be generated by complements and countable sums (see 11.1). As mentioned before, l_x respects these operations, so

$$\nu(x \oplus A^c) = \nu\Big((x \oplus A)^c\Big) = 1 - \nu(x \oplus A) = 1 - \nu(A) = \nu(A^c)$$

and

$$\nu\left(x \oplus \bigcup_i A_i\right) = \nu\left(\bigcup_i x \oplus A_i\right) = \sum_i \nu(x \oplus A_i) = \sum_i \nu(A_i) = \nu\left(\bigcup_i A_i\right).$$

Problem 7

The identification is

$$A \mapsto \chi_A,\; x \mapsto \{n : x_n = 1\}.$$

One easily checks that these two are mutually inverse. Addition modulo 2 comes out to 1 iff exactly one of the summands is 1, and this corresponds exactly to belonging to the symmetric difference.

Problem 8

A filter cannot contain both A and A^c, since then it would contain $A \cap A^c = \varnothing$. Thus, a filter containing for all A either A or A^c is maximal.

For the other direction, suppose neither A nor A^c is in a filter $\mathcal{F}$. We define its *extension* by A as

$$\mathcal{F}_A = \{A' \cap F : A \subseteq A', F \in \mathcal{F}\}.$$

We check that this is a filter.

1. If $\varnothing \in \mathcal{F}_A$, $\mathcal{F}$ contains a set disjoint with A, so by the superset property it contains A^c.
2. Let $A_1 \cap F_1, A_2 \cap F_2 \in \mathcal{F}_A$. Then

   $$A \subseteq A_1 \cap A_2, F_1 \cap F_2 \in \mathcal{F},$$

   so $(A_1 \cap F_1) \cap (A_2 \cap F_2) = (A_1 \cap A_2) \cap (F_1 \cap F_2) \in \mathcal{F}_A$.
3. Let $B \supseteq A' \cap F$. Then

   $$B = B \cup (A' \cap F) = (B \cup A') \cap (B \cup F)$$

   and $A \subseteq A' \cup B$, $F \subseteq B \cup F$, so $B \in \mathcal{F}_A$.

Of course, $A \in \mathcal{F}_A \setminus \mathcal{F}$, so $\mathcal{F}$ was not maximal in the first place.

Remark. One can check that $\mathcal{F}_A$ is the minimal filter containing $\mathcal{F}$ and A.

Problem 9

The only principal ultrafilters are generated by singletons, so they are definitely measurable.

Appendix A

Some values of Tor and Ext

$\mathrm{Tor}_{\mathbb{Z}}$	$\mathbb{Z}$	$\mathbb{Z}_m$	G
$\mathbb{Z}$	0	0	0
$\mathbb{Z}_n$	0	$\mathbb{Z}_{\gcd(m,n)}$	$\ker(G \xrightarrow{\cdot n} G)$