

## Addenda to Course Notes for 447

### 2.5a Other fields

We are mainly interested in the fields of real and complex numbers (and the field of rational numbers). However, the following two fields serve the purpose of better fathoming the scope of the field axioms.

**Example 2.5.x1:** *There is a field with exactly two elements,  $\mathbb{F} = \{e, o\}$ , subject to the rules  $e + e = e$ ,  $e + o = o + e = o$ ,  $o + o = e$ ,  $ee = e$ ,  $eo = oe = e$ , and  $oo = o$ . Checking the field axioms is routine, albeit many cases need to be considered.  $e$  plays the role of 0, and  $o$  plays the role of 1. In particular, this example shows that you cannot prove  $1 + 1 \neq 0$  from the field axioms alone. We can use the order axioms on top of the field axioms to prove that  $1 + 1 \neq 0$  in any ordered field. — The motivation behind this example is that  $o$  stands for ‘odd integer’ and  $e$  stands for ‘even integer’, and the definitions in this field incorporate properties of arithmetic of integers. However, this is just the motivational part, and of course you may not translate the property  $o^{-1} = o$  in the field into a claim that the reciprocal of an odd integer is an odd integer (an obviously false statement). The meaning of  $^{-1}$  comes from the precise wording of the field axioms, not from the motivation for this field.*

The following example gives an ordered field that does not satisfy the supremum axiom (that alone would be easy, take  $\mathbb{Q}$ ), but also does not satisfy the Archimedean property, which we had derived as a consequence of the supremum axiom. [Incidentally, the archimedean field  $\mathbb{Q}$  shows that the supremum axiom is stronger than the archimedean property, i.e., cannot be proved by assuming the archimedean property as an axiom instead.]

The field in question is one whose elements are familiar to you from elementary calculus; however to prove that it satisfies the properties, we need to rely on results from calculus, which we have not rigorously re-proved yet within our axiomatic setting. Since we are not relying on this example in our logical foundation of calculus, this is no problem. You could quarantine the example away unused for now and restudy it later. But I present it here, because I believe that seeing such an example gives you a better feeling for what the axioms do NOT allow to conclude. Namely, it is not possible to prove the boundedness of  $\mathbb{N}$  from the ordered field axioms alone.

**Example 2.5.x2:**  $\mathbb{R}(x)$  denotes the set of all rational functions of a real variable (defined on  $\mathbb{R} \setminus$  finitely many points). (Rational functions are defined to be quotients of polynomials). Addition and multiplication in the field are the usual addition and multiplication of functions. It is easy to see that the field axioms are satisfied. The constant functions 0 and 1 respectively are the neutrals for addition and multiplication. The set  $\mathbb{N}$  can be interpreted as a subset of  $\mathbb{R}(x)$ , by viewing the natural number  $n$  as the constant function  $n$ . Now  $\mathbb{N}$  can be defined from entirely within the field axioms by repeated addition of the 1-element of the field.

Now we define an order  $\prec$  on  $\mathbb{R}(x)$ . (There are many ways how we could do this, I am just giving one choice.) We say  $f \prec g$  iff there exists  $\varepsilon > 0$  such that  $f(x) < g(x)$  for all  $x \in ]0, \varepsilon[$  (the open interval that is denoted as  $(0, \varepsilon)$  by many other authors).

It is almost trivial to show that the transitivity, the addition and the multiplication axioms are verified in this situation. Checking the trichotomy axiom is a bit trickier: Suppose  $f \not\prec g$  and  $g \not\prec f$ . We must show that then  $f = g$ .

Now  $f \not\prec g$  means: For every  $\varepsilon > 0$ , there exists an  $x \in ]0, \varepsilon[$  such that  $f(x) \not\prec g(x)$ . Likewise  $g \not\prec f$  means: For every  $\varepsilon > 0$ , there exists an  $x \in ]0, \varepsilon[$  such that  $g(x) \not\prec f(x)$ . We can

then inductively construct a sequence  $x_1 > x_2 > x_3 > \dots > 0$  such that  $f(x_1) \leq g(x_1)$ ,  $f(x_2) \geq g(x_2)$ ,  $f(x_3) \leq g(x_3)$ ,  $f(x_4) \geq g(x_4)$ , etc with alternating inequality signs. By the intermediate value theorem, we can find a sequence  $\xi_i$  where  $\xi_i \in [x_{i+1}, x_i]$  such that  $f(\xi_i) = g(\xi_i)$ . Since there are infinitely many distinct  $\xi_i$  in this sequence, and since the equation  $f(\xi_i) = g(\xi_i)$  is equivalent to a polynomial equation by cross-multiplication, which has either only finitely many distinct solutions or else is identically fulfilled, it follows that  $f(x) = g(x)$  for all  $x$ , hence  $f = g$ .

[Remember, at this moment it is not crucial that you master every detail of this proof at a formal level, but rather that you understand the example at a pragmatic level, so that it can inform your intuition about the scope of the axioms.]

Now I claim that this ordered field does not satisfy the supremum axiom. The set  $\mathbb{N}$  (consisting of the constant functions whose value is a positive integer) is bounded above, because  $n < \frac{1}{x}$  for every  $n \in \mathbb{N}$ . For the same reason, the archimedean property is not satisfied: Given 1 and  $\frac{1}{x}$  in  $\mathbb{R}(x)$ , there is no natural number  $n$  such that  $n \cdot 1 > \frac{1}{x}$ . Therefore, from the proof of Thm. 2.5.12, this field cannot satisfy the supremum axiom. We can also see this directly: We use  $0 < 1$ , and  $0 < \frac{1}{2}$  (which is the multiplicative inverse of  $1 + 1$ ); these can be proven from the ordered field axioms (how?). Then we want to argue: If  $f$  is an upper bound for  $\mathbb{N}$ , then  $\frac{1}{2}f$  is also an upper bound for  $\mathbb{N}$ , and  $\frac{1}{2}f < f$ . So there is no smallest upper bound, i.e., no supremum of  $\mathbb{N}$ . Can you fill in the details? The trichotomy proof can be simplified; the sequence construction can be avoided.

### 3.1a: More examples of metric spaces

In all the examples of distances below, the symmetry and positivity are trivially satisfied; I will only comment on the triangle inequality, which is not trivial.

Refer to section 5.9 (Def 5.9.1, Prop. 5.9.3). You may replace the abstract notion ‘vector space’ with any of the specific examples  $\mathbb{R}^n$ ,  $C^0[a, b]$ . Given a norm  $\|\cdot\|$ , we obtain a metric by defining  $d(x, y) := \|x - y\|$ .

**Example 3.1.x1:** On  $\mathbb{R}^n$ , using the notation  $x = (x_1, \dots, x_n)$  for its elements  $x$ , we can define the euclidean distance  $d_2$  by  $d_2(x, y) := \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$ . This is a metric, coming from the norm  $\|x\|_2 := \sqrt{x_1^2 + \dots + x_n^2}$ . The non-trivial part is the triangle inequality, which can be proved from the Cauchy-Schwarz inequality in Linear Algebra. We’ll work out the details later (they are not difficult).

**Example 3.1.x2:** We can also introduce the taxi metric  $d_1$  on  $\mathbb{R}^n$ , which comes from a different norm,  $\|x\|_1 := |x_1| + \dots + |x_n|$ . The name comes from the idea that on a rectangular street grid, the road distance is given by the taxi distance, and is distinct from the euclidean distance (which is as the crow flies). The triangle inequality is an easy consequence of the triangle inequality in  $\mathbb{R}$ .

**Example 3.1.x3:** Another metric, called  $d_\infty$ , on  $\mathbb{R}^n$  comes from the supremum norm  $\|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}$ . Proving the triangle inequality is an easy exercise.

**Exercise 3.1.X1:** In  $\mathbb{R}^2$ , plot the sets  $\{x \mid d(0, x) \leq 1\}$  for the metrics  $d_1$ ,  $d_2$ ,  $d_\infty$  respectively.

**Example 3.1.x4:** All these are special cases of the norm  $\|x\|_p := (|x_1|^p + \dots + |x_n|^p)^{1/p}$  for  $p \geq 1$ . Proof of the triangle inequality in this more general case will be postponed until a more convenient opportunity. (The same formula for  $p < 1$  would give a function  $\|\cdot\|_p$  that violates the triangle inequality.)

**Note:** Even for an analysis that is interested only in  $\mathbb{R}^n$ , the generality of the notion of metric space is useful. We will see later (and easily so) that notions of convergence and continuity can be expressed equivalently in terms of either of these metrics; the freedom to choose other metrics than the euclidean can help to give calculational simplicity in proofs.

**Example 3.1.x5:** This example does NOT come from a norm. Let  $S$  be the unit sphere in  $\mathbb{R}^3$ , i.e.,  $\{x \in \mathbb{R}^3 \mid \|x\|_2 = 1\}$ , and denote by  $\cdot$  the dot product in  $\mathbb{R}^3$ . Then the airplane distance is defined by  $d(x, y) := \arccos(x \cdot y)$ . We'll provide a quick proof of the triangle inequality below, for completeness. But the main focus here is that we have an example of a metric on a curved surface that is clearly a reasonable domain on which to study functions. So this example motivates why the generality of the notion 'metric' is useful.

FYI (and you are free to study or to skip this as preferred): Proof of the triangle inequality for the airplane distance: We use the dot and cross products in  $\mathbb{R}^3$ :

For vectors  $A, B, C \in \mathbb{R}^3$ , the following identity holds:

$$(A \cdot B)(C \cdot C) = (A \cdot C)(B \cdot C) + (A \times C) \cdot (B \times C)$$

It is straightforward, albeit lengthy, to prove this identity in terms of components. If we now assume  $\|A\|, \|B\|, \|C\| = 1$  and let  $a := d(B, C)$ ,  $b := d(A, C)$ ,  $c := d(A, B)$ , this formula becomes

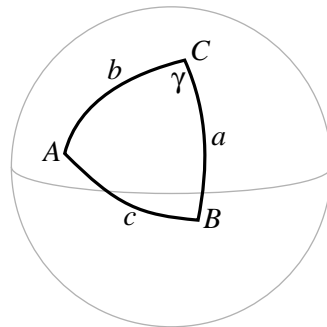
$$\cos c = \cos a \cos b + \sin a \sin b \cos \gamma$$

where  $\gamma$  is the angle between  $B \times C$  and  $A \times C$ . This is called the law of side cosines in spherical trigonometry. (In this context,  $A, B, C$  are vertices of a spherical triangle,  $a, b, c$  are its sides, and  $\gamma$  is the angle at  $C$ ). Now, since  $a, b \in [0, \pi]$  (and so their sines are nonnegative), we conclude

$$\cos c \geq \cos a \cos b - \sin a \sin b = \cos(a + b).$$

If  $a + b \leq \pi$ , we conclude by applying the arccos that  $c \leq a + b$ .

If  $a + b > \pi$ , the inequality  $c \leq a + b$  is trivial.



**Example 3.1.x6:** (This example is interesting in its own right, to get a feel for the generality of the notion of metric, but plays no significant role in calculus. It is a good source for counterexamples to seemingly plausible, but false conjectures.) Choose a prime number  $p$  and define a metric on the set  $\mathbb{Q}$  of rational numbers: Write  $x - y$  as a product of prime powers (e.g.,  $\frac{17}{6} - \frac{8}{9} = \frac{35}{18} = 2^{-1} \cdot 3^{-2} \cdot 5 \cdot 7$ ); in algebra/number theory it is shown that this factorization is unique; the  $p$ -adic distance  $d_p$  from  $x$  to  $y$  is the reciprocal of the power of  $p$  in this factorization. So  $d_2(\frac{17}{6}, \frac{8}{9}) = 2$ ,  $d_3(\frac{17}{6}, \frac{8}{9}) = 9$ ,  $d_5(\frac{17}{6}, \frac{8}{9}) = \frac{1}{5}$ ,  $d_{11}(\frac{17}{6}, \frac{8}{9}) = 1$ . [This  $d_p$  has nothing to do with the  $d_p$ 's on  $\mathbb{R}^n$  mentioned above.] The triangle inequality is satisfied even in the following stronger version:  $d_p(x, z) \leq \max\{d_p(x, y), d_p(y, z)\} \leq d_p(x, y) + d_p(y, z)$ .

**Exercise 3.1.X2:** Show: If  $0 < u \leq v$ , then  $\frac{u}{1+u} \leq \frac{v}{1+v}$ , and use this property to show: If  $d$  is a metric, then  $d_* := d/(1+d)$  is a metric, too.

### 3.3a: lim sup and lim inf

(This material will be covered later – see pg 86. For now, just skim it as a sneak preview.)

**Exercise 3.3X1:** In any metric space, the following holds: The set of cluster points of a sequence is closed.

The trouble with proving convergence of a sequence in  $\mathbb{R}$  is often, that we want to do some calculations involving the presumptive limit of the sequence at a time when we have not proved convergence yet. But this calculation is still meant to be useful in actually proving convergence. The way out of this dilemma is provided by the notions of  $\limsup$  and  $\liminf$ .

**Definition 3.3x1:** Let  $(x_n)$  be a sequence of real numbers. If  $(x_n)$  is NOT bounded above, we define  $\limsup x_n = \infty$ . Otherwise, we define  $\limsup x_n := \lim_{N \rightarrow \infty} \sup_{n \geq N} x_n$ . — Similarly, if  $(x_n)$  is NOT bounded below, we define  $\liminf x_n = -\infty$ . Otherwise, we define  $\liminf x_n := \lim_{N \rightarrow \infty} \inf_{n \geq N} x_n$ .

It is a corollary to Prop. 3.3.15 that the  $\limsup$  and  $\liminf$  of every real sequence exist. It is also immediate that  $\liminf x_n \leq \limsup x_n$

**Exercise 3.3X2:** Show: If  $\lim x_n$  exists then  $\liminf x_n = \lim x_n = \limsup x_n$ . Conversely, if  $\limsup x_n = \liminf x_n$ , then  $\lim x_n$  exists.  $\limsup x_n$  and  $\liminf x_n$  are cluster points of the sequence  $(x_n)$ , and any cluster point  $s$  of the sequence  $x_n$  satisfies  $\liminf x_n \leq s \leq \limsup x_n$ .

The following neat exercise is actually a lemma that is useful in dynamical systems and in ergodic theory. Its proof illustrates the idea to calculate with  $\liminf$  and  $\limsup$  in order to prove later, based on these calculations, that the limit exists.

**Exercise 3.3.X3:** Let  $(x_n)$  be a sequence of real numbers that is *subadditive*, i.e.,  $x_{k+l} \leq x_k + x_l$  for all  $k, l \in \mathbb{N}$ . If  $(x_n)$  is also bounded below, then  $\lim \frac{x_n}{n}$  exists. — Hint: For each  $m, r \in \mathbb{N}$ , consider the sequence  $(x_{km+r}/(km+r))_k$  and show that  $\limsup_{k \rightarrow \infty} \frac{x_{km+r}}{km+r} \leq \frac{x_m}{m}$ . Next conclude that  $\limsup_{n \rightarrow \infty} \frac{x_n}{n} \leq \inf_m \frac{x_m}{m}$ . Finally, conclude the proof. Re-examine your proof. The hypothesis that  $x_n$  be bounded below has not entered in full strength. What weaker hypothesis was actually used?

### 3.4a: Semicontinuity

**Definition 3.4.x1:** A function  $f : X \rightarrow \mathbb{R}$  is called *lower semicontinuous (lsc)* at  $x$ , if for all  $\varepsilon > 0$ , there exists  $\delta > 0$  such that  $f(y) \geq f(x) - \varepsilon$  for all  $y \in B(x, \delta)$ . It is called *upper semicontinuous (usc)* at  $x$  if for all  $\varepsilon > 0$ , there exists  $\delta > 0$  such that  $f(y) \leq f(x) + \varepsilon$  for all  $y \in B(x, \delta)$ .

This definition is occasionally useful. We'll find it helpful during our approach to integration theory. Lower semi-continuity plays a crucial role in minimization problems in calculus of variations, where we ask for the minimum of a function  $f : X \rightarrow \mathbb{R}$  in situations where  $X$  itself is a set consisting of real functions (single or multi-variable).

**Exercise 3.4.X1:** Show:

- (1)  $f : X \rightarrow \mathbb{R}$  is continuous if and only if it is both upper and lower semicontinuous.
- (2) Accepting the properties of the sine from elementary calculus, show that  $f(x) := \sin 1/x$  for  $x \neq 0$  and  $f(0) := a$  defines a function that is lsc at  $x = 0$  iff  $a \leq -1$ , and that is usc at  $x = 0$  iff  $a \geq 1$ .
- (3) Show that  $f : X \rightarrow \mathbb{R}$  is usc iff  $f^{-1}(]-\infty, c])$  is open for every  $c \in \mathbb{R}$ , that  $f : X \rightarrow \mathbb{R}$  is lsc iff  $f^{-1}(]c, \infty])$  is open for every  $c \in \mathbb{R}$ , and that  $f : X \rightarrow \mathbb{R}$  is continuous iff  $f^{-1}(]a, b])$  is open for every  $a, b \in \mathbb{R}$ .
- (4) Show that  $f$  is usc at  $x$  iff  $\lim_{\delta \rightarrow 0} \sup_{B(x, \delta)} f \leq f(x)$ . State an analogous statement for lsc.

### 3.5a: Cover compactness

First observe the remark below definition 3.5.1 and Rmk 3.5.13 in the printed notes: There are two definitions of compactness that are equivalent in metric spaces but become distinct in more general settings. The one presented in the notes according to Def 3.5.1 is usually called ‘sequentially compact’. The other definition (showing up as a conclusion in Prop 3.5.14) is usually just called compact, but may be called ‘cover compact’ for underlining the distinction more clearly.

In addition to the material provided by the book, I’d like to include a full equivalence proof ‘cover compact’  $\iff$  ‘sequentially compact’ in the course. To this end, it may help if you clarify the distinction by replacing the word ‘compact’ with ‘sequentially compact’ in the book all the way up to Exercise 3.37.

Here I will create a clone of the first part of the book’s chapter 3.5, reproving all theorems in terms of ‘cover compact’. This should give you the skill to handle the notion of open covers in practical settings. The cloned theorems will have the corresponding numbers, with a ‘C’ for cover attached.

Read Def 3.5.12.

**Definition 3.5.1C:** *A subset  $A$  of a metric space  $X$  is called cover compact if every open cover  $\{U_\lambda\}_{\lambda \in \Lambda}$  of  $A$  has a finite subcover, i.e., a finite subcollection  $\{A_{\lambda_1}, \dots, A_{\lambda_k}\}$  that covers  $A$ .*

**Proposition 3.5.2C:** *If  $A$  is a cover compact subset of a metric space  $A$ , then  $A$  is closed and bounded.*

**Proof:** There is nothing to prove if  $X = \emptyset$ . For boundedness, we choose  $\Lambda = \mathbb{N}$  and one point  $x_0 \in X$ , and we let  $U_\lambda := B(x_0, \lambda)$ . These provide an open cover for all of  $X$  and therefore for  $A$  also. Since  $A$  is cover compact, there is a finite subcover  $\{U_\lambda \mid \lambda \in \{n_1, \dots, n_k\}\}$ . Without loss of generality, let  $n_k$  be the maximum of  $\{n_1, \dots, n_k\}$ . Then  $A \subset B(x_0, n_k)$ , hence is bounded.

To show that  $A$  is closed, consider  $x \in \bar{A}$ . We want to show  $x \in A$ . To this end, we assume  $x \notin A$ , and from this we construct an open cover of  $A$  that has no finite subcover. Letting  $\Lambda := \mathbb{N}$  again, we define  $U_n := X \setminus C(x, \frac{1}{n})$ , the complement of a closed ball. Since  $\bigcap C(x, \frac{1}{n}) = \{x\}$ , we have  $\bigcup U_n = X \setminus \{x\} \supset A$ . For any finite subcollection of  $\{U_n\}$ , the union would be  $X \setminus C(x, \frac{1}{n_{max}})$ . This does not contain  $A$ , because  $C(x, \frac{1}{n_{max}}) \cap A \supset B(x, \frac{1}{n_{max}}) \cap A \neq \emptyset$ . ■

**Prop. 3.5.3C:** *If  $A$  is a cover compact subset of a metric space  $X$  and  $B$  is a closed subset of  $A$ , then  $B$  is cover compact.*

**Proof:** Let  $\{U_\lambda \mid \lambda \in \Lambda\}$  be an open cover of  $B$ . Then  $\{U_\lambda \mid \lambda \in \Lambda\} \cup \{X \setminus B\}$  is an open cover of  $X$  and therefore of  $A$ . Take a finite subcollection covering  $A$ . It is no loss of generality to assume that  $X \setminus B$  is among this finite subcover, for otherwise we just join it. So we have this finite cover of  $B$  trivially:  $U_{\lambda_1} \cup \dots \cup U_{\lambda_k} \cup (X \setminus B) \supset A \supset B$ . But then  $U_{\lambda_1} \cup \dots \cup U_{\lambda_k} \supset B$  as well, because  $X \setminus B$  is disjoint from  $B$ . ■

**Exercise 3.36C:** Show that any singleton set  $\{x\}$  is cover compact.

Answer (with the usual notation): If  $\{x\} \subset \bigcup U_\lambda$ , then  $x \in \bigcup U_\lambda$ ; hence  $x \in U_{\lambda_0}$  for some  $\lambda_0 \in \Lambda$ , so  $\{x\}$  is covered by a single set, which is a finite subcover.

**Note:** Here and in the following we freely use results like  $\lim 1/(2^n) \rightarrow 0$ . This follows from  $2^i > i$ , which in turn is proved by induction, and the sandwich theorem (Exercise 3.20).

**Thm 3.5.5C:** (Heine-Borel) *A subset of the reals is cover compact if and only if it is bounded and closed.*

**Proof:** The ‘ $\implies$ ’ part is Prop. 3.5.2C. For the converse, assume  $A$  is closed and bounded. Letting  $a := \inf A = \min A$  and  $b := \sup A = \max A$ , we have  $A \subset [a, b]$ . In view of Prop 3.5.3C, it suffices to show that  $[a, b]$  is cover compact. We provide an indirect proof, assuming  $\{U_\lambda\}$  is an open cover of  $[a, b]$  for which there is no finite subcover. Let  $a_0 := a$  and  $b_0 := b$  and construct a sequence of intervals  $[a_i, b_i]$  inductively by successive bisection. Namely, assuming that  $[a_i, b_i]$  has been constructed in such a way that no finite subcollection of  $\{U_\lambda\}$  covers  $[a_i, b_i]$ , we notice that at least one of the following is true: No finite subcollection covers  $[a_i, \frac{1}{2}(a_i + b_i)]$ , or no finite subcollection covers  $[\frac{1}{2}(a_i + b_i), b_i]$ . Accordingly we define  $[a_{i+1}, b_{i+1}]$  as one of these two intervals. Now we have a sequence of nested intervals  $[a_i, b_i] \supset [a_{i+1}, b_{i+1}]$ . The increasing sequence  $(a_i)$ , which is bounded above by  $b_0$ , has a limit, which we call  $x_*$ . The decreasing sequence  $(b_i)$ , which is bounded below by  $a_0$ , also has a limit, which we call  $x^*$ . But by construction  $b_i - a_i = (b_0 - a_0)/2^i \rightarrow 0$ , so  $x_* = x^*$ . Since  $x_* \in [a, b]$ , there must be some  $\lambda$  such that  $U_\lambda \ni x_*$ . Since  $U_\lambda$  is open, there exists  $\varepsilon > 0$  such that  $]x_* - \varepsilon, x_* + \varepsilon[ \subset U_\lambda$ . But as soon as  $(b_0 - a_0)/2^i < \varepsilon$ , which is the case for  $i$  sufficiently large, we will have  $[a_i, b_i] \subset ]x_* - \varepsilon, x_* + \varepsilon[$  (using  $x_* \in [a_i, b_i]$ ). But this contradicts the construction, according to which  $[a_i, b_i]$  cannot be covered by finitely many of the  $U_\lambda$ , specifically not by a single one. ■

**Thm 3.5.6C:** *Let  $X$  and  $Y$  be metric spaces,  $A$  be a cover compact subset of  $X$ , and  $f : X \rightarrow Y$  continuous. Then  $f(A)$  is a cover compact subset of  $Y$ .*

**Proof:** Let  $\{U_\lambda\}$  be an open cover of  $f(A)$ . Then  $\{V_\lambda := f^{-1}(U_\lambda)\}$  is a collection of open sets by Prop. 3.4.8, and it covers  $f^{-1}(f(A))$ . [Can you fill in the details here? Let me just do it one more time, to play it safe. If  $x \in f^{-1}(f(A))$ , this means that  $f(x) \in f(A)$  by definition of inverse image. Since  $\bigcup U_\lambda \supset f(A) \ni f(x)$ , there is some  $\lambda$  such that  $f(x) \in U_\lambda$ . But this means that  $x \in f^{-1}(U_\lambda) = V_\lambda$ .]

Since  $A \subset f^{-1}(f(A))$ , the  $V_\lambda$  cover  $A$ , and since  $A$  is compact, there is a finite subcover  $\{V_{\lambda_1}, \dots, V_{\lambda_k}\}$  for  $A$ . Then the corresponding  $\{U_{\lambda_1}, \dots, U_{\lambda_k}\}$  covers  $f(A)$ . ■

**Thm 3.5.9C:** *Let  $\{A_i\}_{i=1}^\infty$  be a collection of non-empty cover compact subsets of a metric space  $X$  that is nested, in the sense that if  $i < j$  then  $A_j \subset A_i$ . Then  $\bigcap_{i=1}^\infty A_i \neq \emptyset$ .*

**Proof:** Assuming the intersection to be empty, we construct an open cover of  $A_1$  that has no finite subcover. Namely we let  $U_i := X \setminus A_i$ , which are open sets, because the (cover compact)  $A_i$  are closed. The collection  $\{U_i\}$  covers all of  $X$  when  $\bigcap A_i = \emptyset$ . So in particular it covers  $A_1$ . If we consider any finite subcover  $U_{i_1} \cup \dots \cup U_{i_k}$  (which equals  $U_{i_k} = X \setminus A_{i_k}$  if  $i_k$  is the largest of the subscripts), and select some  $x \in A_{i_k} \subset A_1$  (since the  $A$  are non-empty), then we observe that  $U_{i_k} = X \setminus A_{i_k} \not\ni x$ . So this finite collection fails to cover  $A$ . ■

After developing cover-compactness in analogy to sequential compactness, we now prove the equivalence of the two notions in metric spaces.

**Theorem 3.5.14C(1):** *If a metric space  $X$  is sequentially compact, then it is cover compact.*

**Proof:** The case  $X = \emptyset$  being trivial, we now assume  $X \neq \emptyset$ . Let  $\{U_\lambda\}$  be an open cover of  $X$ . By proposition 3.5.14, there exists  $\varepsilon > 0$  such that for every  $x \in X$ , there exists some  $\lambda = \lambda(x)$  such that the ball  $B(x, \varepsilon)$  is entirely contained in  $U_\lambda$ . It suffices to show that the open cover  $\mathcal{B} := \{B(x, \varepsilon)\}_{x \in X}$  has a finite subcover  $\{B(x_1, \varepsilon), \dots, B(x_n, \varepsilon)\}$ . For in this case,  $\{U_{\lambda(x_1)}, \dots, U_{\lambda(x_n)}\}$  will be a finite subcover of the original cover.

Assume  $\mathcal{B}$  does not have a finite subcover. We will construct a sequence that cannot have a convergent subsequence (and hence not a cluster point), in violation of sequential compactness. To this end choose  $y_1 \in X$  arbitrarily. Assuming we have  $y_1, \dots, y_n$  chosen already, we choose  $y_{n+1} \in X \setminus \bigcup_{i=1}^n B(y_i, \varepsilon)$ . This latter set is not empty, because we assumed  $X$  is not covered by any finite subcover of  $\mathcal{B}$ . By construction  $d(y_i, y_j) \geq \varepsilon$  for  $j > i$ . And by symmetry the same holds for  $j \neq i$ . If the sequence  $(y_i)$  had a convergent subsequence, whose limit may be called  $y_*$ , then infinitely many of the  $y_i$  would have to be within  $B(y_*, \frac{1}{2}\varepsilon)$ , but this contradicts the fact that any two of the  $y_i$  have distance  $\geq \varepsilon$ . ■

**Theorem 3.5.14C(2):** *If a metric space  $X$  is cover compact, then it is sequentially compact.*

**Proof:** by contrapositive. Suppose  $(x_n)_{n=1}^\infty$  is a sequence without cluster point. We construct an open cover of  $X$  that has no finite subcover. The set  $A := \{x_n\}$  is closed by Prop. 3.3.12 (because the sequence has no cluster point), so  $U_0 := X \setminus A$  is open. Since the sequence  $(x_n)$  has no cluster point, in particular  $x_1$  is not a cluster point, so there exists  $\varepsilon_1 > 0$  such that  $B(x_1, \varepsilon_1)$  contains at most finitely many terms of the sequence. Disregarding those that may be equal to  $x_1$ , we can decrease  $\varepsilon_1$  to make sure that  $B(x_1, \varepsilon_1)$  contains no other terms of the sequence (except possible repeats, of which there are at most finitely many). In other words  $B(x_1, \varepsilon_1)$  is disjoint from the set  $A \setminus \{x_1\}$ . Inductively, once we have constructed  $x_1, \dots, x_n$  and  $\varepsilon_1 \geq \dots \geq \varepsilon_n$ , we find  $x_{n+1}$  and  $\varepsilon_{n+1} \leq \varepsilon_n$  such that  $B(x_{n+1}, \varepsilon_{n+1})$  is disjoint from  $A \setminus \{x_1, \dots, x_n\}$ . Now the balls  $U_n := B(x_n, \frac{1}{2}\varepsilon_n)$  are pairwise disjoint, because  $y \in B(x_n, \frac{1}{2}\varepsilon_n) \cap B(x_m, \frac{1}{2}\varepsilon_m)$  (with  $m > n$ ) would imply  $d(x_n, x_m) \leq d(x_n, y) + d(y, x_m) < \frac{1}{2}\varepsilon_n + \frac{1}{2}\varepsilon_m \leq \varepsilon_n$ , so  $x_m$  would be in  $B(x_n, \varepsilon_n)$ , contrary to the construction.

Now  $\bigcup_{n=0}^\infty U_n = X$ , but no finite subcollection (without loss of generality  $\{U_n\}_{n=0}^N$ ) can cover  $X$ , specifically it does not cover  $x_{N+1}$ . ■

### 3.5b: Cantor sets

An example of a compact set in  $\mathbb{R}$ , in two variants. Every analyst needs to know this example, which is good in killing many a too-naive conjecture.

**Example 3.5.x1:** Let  $A_0 := [0, 1]$ . Let  $A_1 := [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ . Inductively define  $A_n$  as the union of  $2^n$  many closed intervals, each of length  $3^{-n}$ , obtained by removing the middle third from each of the intervals of which  $A_{n-1}$  is made up. The standard Cantor set is defined as  $C := \bigcap A_n$ .

As an intersection of a nested sequence of compact non-empty sets, it is non-empty and compact (by Thm 3.5.9). In the spirit of Cor 3.5.10 (with a few details pertaining to the next sections on subspaces skipped here), it can be shown that  $C$  is uncountable.

While we have yet to define rigorously a notion of measure (generalizing the notion of length of an interval in  $\mathbb{R}$ ), let us just notice informally with the purpose of motivating a future definition of length ('measure'), that any sensible length we may assign to  $C$  ought to be less

that  $2^n/3^n$ , because  $C \subset A_n$ . So, being nonnegative by another reasonable stipulation, the length ('measure') of  $C$  ought to be (and will be) 0.

$C$  is 'full of holes' in the following sense: Between any two points  $x, y \in C$  with  $x < y$ , there exists a point  $z$  that is NOT in  $C$ . In particular,  $\overset{\circ}{C} = \emptyset$ .

**Example 3.5.x2:** 'Fat Cantor set'. This is a variant of the same construction. The fat Cantor set will share all properties of the Cantor set that were mentioned above, except that it 'ought to have' (and will have, once rigorously defined) positive length (positive measure). Again we start with  $A_0 = [0, 1]$ . But now  $A_n$  is designed to have length  $\frac{1}{2} + 2^{-n-1}$ . To this end, remove an interval of length  $\frac{1}{4}$  from the middle of  $A_0$  to get  $A_1 = [0, \frac{3}{8}] \cup [\frac{5}{8}, 1]$ . Now remove two intervals of total length  $\frac{1}{8}$  (namely length  $\frac{1}{16}$  each) from the middle of each of the two intervals that make up  $A_1$ . This gives us  $A_2 = [0, \frac{5}{32}] \cup [\frac{7}{32}, \frac{3}{8}] \cup [\frac{5}{8}, \frac{25}{32}] \cup [\frac{27}{32}, 1]$ . Continuing this way, let  $C_{\text{fat}} := \bigcap A_n$ .

The length of  $C_{\text{fat}}$  ought to be (and will be defined as)  $\lim(\frac{1}{2} + 2^{-n-1}) = \frac{1}{2}$ .

**Addendum to Example 3.5.11:** In the space  $X := C^0[0, 1]$  of continuous real valued functions on  $[0, 1]$ , with the sup distance, here is an example of a nested family  $A_n$  of bounded and closed sets, with empty intersection. Note that in this metric space, bounded and closed does not imply compact. (At least we haven't proved any theorem to this effect, and this example will now show that such an implication can indeed not hold.)

Take  $A_n := \{f \in C^0[0, 1] \mid f(0) = 0 \text{ and } \frac{nx}{1+nx} \leq f(x) \leq 1 \text{ for all } x \in ]0, 1]\}$ .

The sets  $A_n$  are clearly bounded, since they lie in the ball  $B(\text{zerofunction}, 1.001)$ . The nesting of the sequence is trivial. To see that the sets  $A_n$  are closed, we may notice that they are defined by pointwise non-strict inequalities ( $\leq$ ), and these automatically define *closed* sets in  $C^0[0, 1]$ . Indeed, consider, for each  $x \in ]0, 1]$ , the function  $\text{ev}_x : C^0[0, 1] \rightarrow \mathbb{R}$ , defined by  $\text{ev}_x(f) := f(x)$ . These are called the evaluation functions. They are continuous, with  $\delta = \varepsilon$ , because  $|f(x) - g(x)| \leq d(f, g)$ . *If you find this confusing, make sure you distinguish carefully between  $f \in C^0[0, 1]$ , which is a function, and its values  $f(x) \in \mathbb{R}$ , which are real numbers, and the function  $\text{ev}_x$ , which assigns to the function  $f$  its value at a given  $x$ .*

So  $A_n = \text{ev}_0^{-1}(\{0\}) \cap \bigcap_{x \in ]0, 1]} \text{ev}_x^{-1}([\frac{nx}{1+nx}, 1])$ . Inverse images of closed sets under continuous functions are closed, and therefore  $A_n$  is closed as an intersection of closed sets.

Now any continuous function  $f \in \bigcap A_n$  must satisfy  $f(0) = 0$  and  $\frac{nx}{1+nx} \leq f(x) \leq 1$  for every  $x > 0$  and every  $n \in \mathbb{N}$ . But since  $\sup_n \frac{nx}{1+nx} = \lim_{n \rightarrow \infty} \frac{nx}{1+nx} = 1$  for every  $x > 0$ , this implies  $f(x) = 1$  for  $x > 0$ , whereas  $f(0) = 0$ . This contradicts continuity of  $f$ .

### 3.9a Examples for the failure of a Heine-Borel type theorem in metric spaces other than $\mathbb{R}^n$ :

In  $\mathbb{Q}$ , there is a bounded sequence that has no convergent subsequence: For instance the sequence  $x_1 = 1, x_2 = \frac{3}{2}, x_3 = \frac{17}{12}, \dots$ , which is recursively defined by  $x_{n+1} := \frac{1}{2}(x_n + 2/x_n)$  defines a bounded sequence in  $\mathbb{Q}$ , which, considered as a sequence in  $\mathbb{R}$ , has the limit  $\sqrt{2}$  (as can be proved with a little labor, but is not the main issue here). Considered as a sequence in  $\mathbb{Q}$ , it does not have a limit in  $\mathbb{Q}$ , nor does any subsequence.

In a discrete metric space  $X$ , *all sequences are bounded*. Now if we take a discrete metric space with infinitely many elements, we can construct a sequence that has no repeat elements (i.e., a function  $\mathbb{N} \rightarrow X$  that is 1-1). Such a sequence has no cluster point.

The following examples are more interesting:

**Hwk 3.9X1:** Let  $X = C^0[-1, 1]$  with the max distance. Define  $f_n \in X$  by  $f_n(x) := nx/\sqrt{1 + (nx)^2}$ . The sequence  $(f_n)$  is bounded, since it lies in  $B(\text{zerofunction}, 1)$ . But prove that it does not have a convergent subsequence. *Hint: Assume some subsequence does have a limit  $f$ . Use the evaluation functions  $\text{ev}_x : C^0[0, 1] \rightarrow \mathbb{R}$  defined by  $\text{ev}_x(f) := f(x)$  to find out what  $f$  would have to be and obtain a contradiction.*

**Hwk 3.9X2:** Let  $X = C^0[-1, 1]$  with the max distance. Define  $f_n \in X$  by  $f_n(x) := nx/(1 + (nx)^2)$ . The sequence  $(f_n)$  is bounded, since it lies in  $B(\text{zerofunction}, 1)$ . But prove that it does not have a convergent subsequence. *Hint: Same as before; but note that the contradiction to be obtained is of a different kind this time. You may want to calculate  $d(f_n, 0)$ .*

**Hwk 3.9X3:** Let  $X = C^0[-1, 1]$  with the max distance. Define  $f_n \in X$  by  $f_n(x) := \sin nx$ . The sequence  $(f_n)$  is bounded again. But prove that it does not have a convergent subsequence. *Hint: Assume there is a convergent subsequence with limit  $f$ . Use uniform continuity of  $f$  to prove:*

$$\forall \varepsilon > 0 \exists \delta > 0 \forall g \in B(f, \varepsilon) \sup_{[-\delta, \delta]} g - \inf_{[-\delta, \delta]} g < 3\varepsilon \quad (EC)$$

*Note that  $\delta$  must not depend on  $g$  (but may depend on  $f$ ). Derive a contradiction by having  $f_n \notin B(f, \varepsilon)$ .*

### 3.10a: Two problems about Connectedness

**Example and comment:**  $\mathbb{R}$  is not homeomorphic to  $\mathbb{R}^n$  for any  $n > 1$ , because the complement of a singleton in  $\mathbb{R}$  is not connected whereas the complement of a singleton in  $\mathbb{R}^n$  is connected, as can be easily verified. — It is true, but much more sophisticated to prove that  $\mathbb{R}^n$  is not homeomorphic to  $\mathbb{R}^m$ , except if  $n = m$ . The tools to prove this theorem belong in the area of algebraic topology, and they are beyond 447/448. — Another famous theorem in this area is the Jordan curve theorem: If  $C$  is a simple closed curve in  $\mathbb{R}^2$ , then its complement is not connected, but is the union of exactly two connected subsets of  $\mathbb{R}^2$ . Here, ‘simple closed curve’ means the image of a continuous and injective function from  $\{x \in \mathbb{R}^2 \mid \|x\| = 1\}$  to  $\mathbb{R}^2$ . This theorem, intuitively plausible, is also quite sophisticated. Regardless of whether you ever study a proof of these results in a class or not, consider the contents of this paragraph as core part of the GenEd of a mathematician.

**Hwk 3.10.X1:** Show that the only connected subsets of a Cantor set are singletons.

**Hwk 3.10.X2:** Is  $\mathbb{R}^2 \setminus \mathbb{Q}^2$  connected or not? Explain.

### 3.11a: Banach’s Fixed Point Theorem:

The following theorem is the simplest among a collection of power tools of advanced calculus. Its applications include the Newton method for solving systems of equations, the local existence and uniqueness theorem for ordinary differential equations, and similar results for a variety of partial differential equations.

**Theorem:** (Banach’s fixed point theorem, aka contraction mapping principle) *Suppose  $X \neq \emptyset$  is a complete metric space and  $f : X \rightarrow X$  is a contraction, i.e., it satisfies  $d(f(x), f(y)) \leq \vartheta d(x, y)$  for some constant  $\vartheta < 1$ . Then, in  $X$ , there exists exactly one solution  $x$  to the equation  $f(x) = x$ .*

**Proof idea:** Choose any  $x_0$  in  $X$ . Define recursively  $x_n := f(x_{n-1})$ . Show that  $(x_n)$  is a Cauchy sequence, by getting a recursive estimate for  $d(x_k, x_{k-1})$  and using the triangle inequality. Its limit solves the equation. Finally prove uniqueness.

**Experiment:** As an application for  $X = [0, 1]$ , you can solve the equation  $\cos x = x$  by repeatedly hitting the cos key on your pocket calculator.

**Tourism Corollary for  $X = \text{Knoxville}$ :** If you lay down a city map of Knoxville on the ground anywhere within Knoxville, then there is exactly one point in Knoxville that lies right beneath its corresponding representation on the map. This is true even if the map is (partially) folded up, lying face down or is of the special patented kind, which introduces a distortion to the effect of having a larger scale for the center than for the suburbs.

**Hwk 3.11X1:** Work out the proof details based on the above idea.

**Hwk 3.11X2:** Give an example for  $X = \mathbb{R}$  to the effect that BFPT becomes false if the hypothesis ' $d(f(x), f(y)) \leq \vartheta d(x, y)$  for some constant  $\vartheta < 1$ ' is replaced by the weaker hypothesis  $d(f(x), f(y)) < d(x, y)$  whenever  $x \neq y$ .

**Hwk 3.11X3:** Show that the (false) BFPT from the previous exercise becomes true again, if we throw in the extra hypothesis that  $X$  is compact. Namely show: If  $X$  is a compact metric space, and  $f : X \rightarrow X$  satisfies  $d(f(x), f(y)) < d(x, y)$  whenever  $x \neq y$ , then there exists exactly one solution  $x$  to the equation  $x = f(x)$ . *Hint: Consider the sets  $X_0 := X$ ,  $X_n := f(X_{n-1})$  and study the intersection  $K := \bigcap X_n$ . In particular show  $f(K) = K$ .*

**Comment:** The last two exercises serve the purpose of better understanding the first; in stark contrast to BFPT, they seem to have little application themselves.

**Theorem:**  $C^0[a, b]$ , together with the max distance is a complete metric space.

This is easy to prove (and for that matter generalizes to  $C^0(K)$  where  $K$  is any compact subset of  $\mathbb{R}^n$ , and with minor modifications to yet vaster generality):

**Proof:** Let  $(f_n)$  be a Cauchy sequence in  $C^0[a, b]$ :

$$\forall \varepsilon > 0 \exists n_0 \forall m, n \geq n_0 : \max_{x \in [a, b]} |f_n(x) - f_m(x)| = d(f_n, f_m) < \varepsilon$$

It follows that for each  $x \in [a, b]$ , the sequence  $(f_n(x))_n$  is a Cauchy sequence in  $\mathbb{R}$ . Since  $\mathbb{R}$  is complete, this Cauchy sequence has a limit, and we call it  $f(x)$ . In other words, we have defined a function  $f : [a, b] \rightarrow \mathbb{R}$  that assigns to each  $x \in [a, b]$  the limit  $\lim_{n \rightarrow \infty} f_n(x)$ . We next aim to prove that this function is indeed continuous and hence lies in  $C^0[a, b]$ . After that we will show that  $d(f_n, f) \rightarrow 0$ , i.e., that  $f$  is indeed the limit of the Cauchy sequence  $(f_n)$ .

First the proof that  $f$  is continuous at each  $x$ : Let  $\varepsilon > 0$  and choose  $n_0$  so large that  $d(f_n, f_m) < \frac{\varepsilon}{3}$  for all  $m, n \geq n_0$ . Then  $|f_n(t) - f_m(t)| < \frac{\varepsilon}{3}$  for all  $t$ . Now estimate

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f_m(x)| + |f_m(x) - f_m(y)| + |f_m(y) - f_n(y)| < \frac{2}{3}\varepsilon + |f_m(x) - f_m(y)|$$

We may let  $n \rightarrow \infty$  in this estimate, since we know already that the sequence of real numbers  $f_n(t)$  converges for every  $t$ ; and we get

$$|f(x) - f(y)| \leq \frac{2}{3}\varepsilon + |f_m(x) - f_m(y)|$$

This is true for every fixed  $m \geq n_0(\varepsilon)$ , for instance we can just take  $m = n_0$ . Since this  $f_m$  is continuous, we can find a  $\delta > 0$  such that  $|x - y| < \delta$  implies  $|f_m(x) - f_m(y)| < \frac{\varepsilon}{3}$ . With this choice of  $\delta$ , we then get immediately that  $|f(x) - f(y)| < \varepsilon$ , provided  $|x - y| < \delta$ . We have showed the continuity of  $f$ .

Next we want to show  $d(f_n, f) \rightarrow 0$ . We rewrite the Cauchy sequence property:

$$\forall \varepsilon > 0 \exists n_0 \forall m, n \geq n_0 \forall x \in [a, b] : |f_n(x) - f_m(x)| < 0.9\varepsilon$$

Taking the limit  $n \rightarrow \infty$  in this estimate, we get

$$\forall \varepsilon > 0 \exists n_0 \forall m \geq n_0 \forall x \in [a, b] : |f(x) - f_m(x)| \leq 0.9\varepsilon < \varepsilon$$

But this means  $d(f, f_m) \leq 0.9\varepsilon < \varepsilon$ , and the theorem is proved.

**Example:**  $C^0[a, b]$  can also be equipped with the metric  $d_1(f, g) := \int_a^b |f(x) - g(x)| dx$ . With this metric, the space is NOT complete.

**Hwk 3.11X4:** Show that the sequences in homeworks 3.9.X1 and 3.9.X2 are Cauchy sequences with respect to the metric  $d_1$ . *Hint:* The second sequence is easier, b/c you can just estimate  $\int |f - g| \leq \int |f| + \int |g|$ , and it will be good enough whereas in the first example you should use more care in estimating the difference. Obviously you are allowed to use elementary calculus knowledge here to handle the integrals.

### Compact sets in $C^0[a, b]$

Have another look at condition (EC) in the hint for problem 3.9.X3. The punchline was that the modulus of continuity (i.e., the  $\delta$  chosen as a function of  $\varepsilon$  such that the definition of continuity is satisfied for this choice of  $\delta$ ) was the same for all functions  $g$  in the ball.

**Definition:** A set  $S \subset C^0[a, b]$  is called equicontinuous iff

$$\forall \varepsilon > 0 \exists \delta > 0 \forall g \in S : |x - y| < \delta \implies |g(x) - g(y)| < \varepsilon$$

This definition includes uniform continuity, which is anyways automatically satisfied, because  $[a, b]$  is compact. But the crucial issue is that  $\delta$  does *not* depend on  $g \in S$  (because the  $\forall g$  quantifier follows the  $\exists \delta$  quantifier, rather than preceding it).

Here is a famous theorem that characterizes compact sets in  $C^0[a, b]$  in a similar way as Heine Borel characterizes compact sets in  $\mathbb{R}^n$ :

**Theorem:** (Arzelà-Ascoli) A set  $S \subset C^0[a, b]$  is compact if and only if it is closed, bounded, and equicontinuous.

We'll see later if we get around proving this theorem. I think it is desirable to include it in the course, but it is not required core material. Key proof ingredients are Heine-Borel for the sets  $ev_x(S) \subset \mathbb{R}$ , selection of a countable dense subset of  $[a, b]$  from which the  $x$ 's will be chosen, and a tricky way of successive selection of subsequences known under the label 'diagonal sequence argument'. The Arzelà-Ascoli theorem is another core power tool of modern analysis, and it spawns a whole family of corollaries for other metric spaces consisting of functions.

## 4.2a: AGM inequality

A useful tool for estimates is sometimes the following

**4.2.x1 Theorem:** (Inequality of the arithmetic and geometric mean, in short agm inequality)

Let  $a_1, \dots, a_n$  be nonnegative real numbers. Then

$$\sqrt[n]{a_1 \cdots a_n} \leq \frac{a_1 + \cdots + a_n}{n}$$

Equality holds if and only if  $a_1 = \cdots = a_n$ .

**Note:** The left side of this inequality is called the geometric mean of the  $a_i$ , and the right side is called their arithmetic mean.

**Proof:** We may assume  $a_i > 0$ , b/c the inequality becomes trivial when one of the terms = 0. The case  $n = 1$  is trivial.

We first prove the case  $n = 2$ . This will be the start of an induction over  $k$  that proves the cases  $n = 2^k$ . In a third step, we deal with the other  $n$  by means of an interpolation technique.

For  $n = 2$ , note that  $\sqrt{a_1 a_2} \leq \frac{1}{2}(a_1 + a_2)$  is equivalent to  $4a_1 a_2 \leq (a_1 + a_2)^2$ , and this in turn is equivalent to  $0 \leq (a_1 - a_2)^2$ . Unless  $a_1 = a_2$ , the inequalities are strict. This proves the case  $n = 2$ .

Now assume we have proved the agm inequality for  $n = 2^k$  with some  $k \geq 1$ . We conclude it for  $n' = 2^{k+1} = 2n$  as follows (subscripts to  $\leq$  explain why the inequality holds):

$$\begin{aligned} \sqrt[2n]{a_1 \cdots a_n a_{n+1} \cdots a_{2n}} &= \sqrt{\sqrt[n]{a_1 \cdots a_n} \sqrt[n]{a_{n+1} \cdots a_{2n}}} \leq_{\text{agm } 2} \frac{\sqrt[n]{a_1 \cdots a_n} + \sqrt[n]{a_{n+1} \cdots a_{2n}}}{2} \\ &\leq_{\text{agm } n} \frac{\frac{a_1 + \cdots + a_n}{n} + \frac{a_{n+1} + \cdots + a_{2n}}{n}}{2} = \frac{a_1 + \cdots + a_{2n}}{2n} \end{aligned}$$

For equality to hold, both  $\leq$  signs must be =, and by induction hypothesis used in the second inequality, this requires  $a_1 = \cdots = a_n$ , which quantity we call  $A$ , and  $a_{n+1} = \cdots = a_{2n}$ , which we call  $B$ . We also need  $\sqrt[n]{a_1 \cdots a_n} = \sqrt[n]{a_{n+1} \cdots a_{2n}}$  from the first inequality, so  $A = B$ . By induction the cases  $n = 2^k$  are proved.

Finally, take  $3 \leq n \in \mathbb{N}$  arbitrary. Find an  $N = 2^k$  such that  $N > n$ , and let  $m = N - n$ . We 'pad' the list  $a_1, \dots, a_n$  with  $m$  copies of the arithmetic mean:  $a_j := A := \frac{1}{n}(a_1 + \cdots + a_n)$  for  $j = n + 1, \dots, N$ . Then agm for  $N$  says:

$$(a_1 \cdots a_n A^m)^{1/(n+m)} \leq \frac{a_1 + \cdots + a_n + mA}{n+m} = A$$

The inequality is strict unless  $a_1 = \cdots = a_n = A$ . Cancelling the factor  $A^{m/(n+m)}$  gives

$$(a_1 \cdots a_n)^{1/(n+m)} \leq A^{1 - \frac{m}{n+m}} = A^{n/(n+m)}$$

Taking this to the power  $(n+m)/n$  proves agm for  $n$ . ■

The agm inequality gives an alternate way of proving Exercise 4.12. Namely, applying agm to  $k$ ,  $n/k$ , and  $n - 2$  copies of 1, gives

$$1 \leq \sqrt[n]{n} = \sqrt[n]{k \frac{n}{k} 1^{n-2}} \leq_{\text{agm}} \frac{1}{n} \left( k + \frac{n}{k} + (n-2) \right)$$

We may for instance take  $k = \sqrt{n}$  to conclude  $1 \leq \sqrt[n]{n} \leq \frac{1}{n}(n - 2 + 2\sqrt{n}) < 1 + 2n^{-1/2}$ .

**Comment:** The skill to apply such inequalities wisely is non-trivial, but is a core skill for an analyst. Take special notice when you see inequalities applied in this or other courses, thus striving to build this skill.

#### 4.4a: More on convergence properties

It is a quick corollary to Prop 4.3.8 that, when both series  $\sum a_n$  and  $\sum b_n$  are absolutely convergent, then so is their product series.

The alternating series test allows us to construct an example of two conditionally convergent series whose product as defined by (4.4) in the book diverges. Namely take  $\sum_{n=0}^{\infty} (-1)^n / \sqrt{n+1}$  and take the product of this series with itself. The terms  $c_n$  are given by

$$c_n = \sum_{k=0}^n (-1)^k (k+1)^{-1/2} (-1)^{n-k} (n-k+1)^{-1/2} = (-1)^n \sum_{k=0}^n \frac{1}{\sqrt{(k+1)(n+1-k)}}$$

By the agm inequality, we have

$$|c_n| \geq \sum_{k=0}^n \frac{2}{k+1+n+1-k} = \frac{2(n+1)}{n+2} > 1$$

hence  $\sum c_n$  diverges.

In contrast to the rearrangement of absolutely convergent series (Prop. 4.4.9), we have the following theorem:

**Prop. 4.4.9a:** *Assume  $\sum a_n$  is a series of real numbers that converges, but NOT absolutely. Then for every  $s \in \mathbb{R}$ , there is a rearrangement  $\psi : \mathbb{N} \rightarrow \mathbb{N}$  such that  $\sum a_{\psi(n)}$  converges to  $s$ . Moreover, for any interval  $[s, t]$  or  $[s, \infty[$ , or  $]-\infty, t]$ , or  $]-\infty, \infty[$ , there exists a rearrangement  $\psi$  for which the set of cluster points of the sequence  $(\sum_{n=0}^N a_{\psi(n)})_N$  is said interval.*

**Proof:** For any real number  $a$ , denote  $a_+ := \max\{a, 0\}$  and  $a_- := \max\{-a, 0\}$ , so that  $a = a_+ - a_-$  and  $|a| = a_+ + a_-$ . Both  $a_+$  and  $a_-$  are nonnegative.

Then  $\sum (a_n)_-$  and  $\sum (a_n)_+$  both diverge: for if  $\sum (a_n)_-$  were to converge, then so would  $\sum [a_n + (a_n)_-] = \sum (a_n)_+$ ; and then  $\sum [(a_n)_+ + (a_n)_-] = \sum |a_n|$  would also converge, in contradiction to the hypothesis. A similar contradiction would ensue if  $\sum (a_n)_+$  were to converge.

We first show how to get the interval  $[s, t]$  as cluster points: First take as many of the nonnegative terms  $a_n$  (in their natural order) as are needed to make their sum  $\geq t$ . This is possible since  $\sum (a_n)_+$  diverges. These terms will be the first  $a_{\psi(j)}$ . Next take as many of the negative  $a_n$  as are needed to make the total sum  $\leq s$ . These will be the next  $a_{\psi(j)}$ . Now take again as many nonnegative terms to bring the sum back up above  $t$ . This constructs a rearrangement  $\psi$  inductively for which the limsup of the partial sums is  $\geq t$  and the liminf is  $\leq s$ . Actually, since the  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ , the limsup and liminf are exactly  $t$  and  $s$  respectively. Again since the  $a_n \rightarrow 0$ , all points in the interval  $[s, t]$  are cluster points as well.

The choice  $t = s$  gives a rearrangement with limit  $s$ . If we want  $t\infty$  instead, we select first as many nonnegative terms as are needed to bring the sum above 1, in the third step we bring the sum above 2, and so on. ■

## 4.5b: Formal power series

A formal power series is defined to be a sequence of complex numbers  $(a_n)_{n=0}^{\infty}$ , that is written suggestively as  $\sum a_n Z^n$  (with the  $\sum$  and  $Z^n$  symbols not implying any of the operations yet that they seem to imply). Later, when we discuss convergence of formal power series, then these symbols will obtain their ordinary meaning in that context.

The sum of the formal power series  $(a_n)$  (written as  $\sum a_n Z^n$ ) and  $(b_n)$  (written as  $\sum b_n Z^n$ ) is the formal power series  $(a_n + b_n)$ , written as  $\sum (a_n + b_n) Z^n$ .

The product of these same formal power series is  $(c_n)$ , written as  $\sum c_n Z^n$ , where  $c_n := \sum_{k=0}^n a_k b_{n-k}$ .

With these two operations, it is easily seen that formal power series satisfy all field axioms with one exception: not every nonzero formal power series has a multiplicative inverse; only those for which  $a_0 \neq 0$  do. (This deficit could be mended if we allowed to start indexing at negative  $n$ .) Let's see why this is the case:

First, the multiplicative identity is the sequence  $(1, 0, 0, 0, \dots)$ , written as  $1 + \sum_{n=1}^{\infty} 0 Z^n$ , or more briefly, 1. If  $(\sum a_n Z^n)(\sum b_n Z^n) = 1$ , we clearly must have  $a_0 b_0 = 1$  as a necessary condition, and hence  $a_0 \neq 0$ . Conversely, if  $a_0 \neq 0$ , we can calculate recursively  $b_0 = 1/a_0$  and  $b_n := -\sum_{k=1}^n a_k b_{n-k}/a_0$  for  $n = 1, 2, 3, \dots$

Other operations that are common for functions can be defined completely formally (in terms of finite algebra, with no need for a notion of convergence): For instance, we can define the formal derivative of  $\sum a_n Z^n$  as  $\sum (n+1)a_{n+1} Z^n$ , and we could verify the Leibniz product rule algebraically for this algebraic definition of the derivative.

We could even define a composition of formal power series, obtained by plugging in  $\sum b_n W^n$  for  $Z$  in  $\sum a_n Z^n$ , PROVIDED  $b_0 = 0$  (This hypothesis is crucial!). The result would be a series  $\sum c_n W^n$ , where  $c_n$  is a polynomial expression involving only  $a_0, \dots, a_n, b_1, \dots, b_n$ .

Formal power series can serve as an elegant bookkeeping device in combinatorics. However, in analysis, we want to work with convergent power series.

*[At this point, merge into sec 4.5 of the book]*

We have seen that the sum of absolutely convergent series is absolutely convergent, and that the product of absolutely convergent series is absolutely convergent. If  $\sum a_n z^n$  has radius of convergence  $\rho_1$  and  $\sum b_n z^n$  has radius of convergence  $\rho_2$ , then their sum and their product have radius of convergence *at least*  $\min\{\rho_1, \rho_2\}$ . The radius of convergence could be larger than this, basically due to cancellations of large coefficients: For example,

$$(1 + 2z + 2z^2 + 2z^3 + 2z^4 + \dots)(1 - 2z + 2z^2 - 2z^3 + 2z^4 - + \dots) = 1$$

Both factors have radius of convergence 1 (and their respective sums are actually  $1 + 2z \frac{1}{1-z} = \frac{1+z}{1-z}$  and  $1 - 2z \frac{1}{1+z} = \frac{1-z}{1+z}$ ). The product has radius of convergence infinity.

This example gives a glimpse of a deep insight from complex variables, namely that it is possible to obtain the radius of convergence not only from the coefficients according to Hadamard's formula  $\rho = 1/\limsup \sqrt[n]{|a_n|}$ , but also from differentiability properties of the function represented by the power series. In this example, the 'singularities' at  $z = 1$  and  $z = -1$  respectively were the deeper reason that the power series representing the functions  $z \mapsto \frac{1+z}{1-z}$  and  $z \mapsto \frac{1-z}{1+z}$  would not converge beyond the disk  $|z| < 1$ . We will see that we can understand one direction of this observation easily: namely, we will show that the function represented by a power series in its (open) disc of convergence is continuous there (and even differentiable).

So clearly, the power series representing the function  $z \mapsto \frac{1+z}{1-z}$  cannot possibly converge in a larger disc than  $|z| < 1$ . [Or I should rather have said: ‘a’ power series, until uniqueness of such a power series is proved; but uniqueness does hold.] The converse, namely that the radius of convergence is as large as it can possibly be, taking singularities into account, is a masterpiece of complex variables, which we have alas not the time to pursue in this course. Obviously, any proof of such a result would need to be based on a definition of what are these ‘singularities’ that need to be taken into account. This is an area where it becomes crucial that we are in the complex domain, and it has no analog for real variable functions.

For the material that we do cover in this course, every substantial result will apply equally to real and complex variables.

Finally, let us prove a theorem about the multiplicative inverse of convergent over series:

**Theorem:** *Let  $\sum a_n z^n$  have positive radius of convergence. Let  $a_0 \neq 0$ . Then the reciprocal  $\sum b_n z^n$  with  $b_n$  defined recursively as  $b_0 = 1/a_0$  and  $b_n = -\sum_{k=1}^n a_k b_{n-k}/a_0$  has also positive radius of convergence.*

Note that this theorem does not make any predictions about how large the radius of convergence is. This is natural: The truncated power series  $1 - z$  has trivially infinite radius of convergence, but its reciprocal, the geometric series has radius of convergence 1, and it cannot be any larger because of the behavior of  $\frac{1}{1-z}$  as  $z \rightarrow 1$  (if we accept the easy part of the insight connecting radius of convergence with the represented function). The proof will give some insight into this matter.

The proof relies on the following

**Lemma:** *If  $|a_n| \leq Mr^n$  for some constants  $M, r > 0$  and all  $n \geq 0$ , then  $|b_n| \leq M' \rho^n$  for  $n \geq 1$ , with  $M' = \frac{M}{|a_0|(M+|a_0|)}$  and  $\rho = (M + |a_0|)r/|a_0|$ .*

**Proof:** A straightforward induction.

The lemma does not claim anything for  $b_0$ , but we know  $b_0 = 1/a_0$ . The induction starts at  $n = 1$ , with  $b_1 = -a_1 b_0/a_0$ , hence  $|b_1| \leq Mr/|a_0|^2 = M' \rho$ . Let’s study an induction step from  $n$  to  $n + 1$ , for  $n \geq 1$ :

$$\begin{aligned}
|b_{n+1}| &\leq \sum_{k=1}^{n+1} |a_k| |b_{n+1-k}| / |a_0| \leq \sum_{k=1}^n Mr^k M' \rho^{n+1-k} / |a_0| + Mr^{n+1} / |a_0|^2 = \\
&= \sum_{k=1}^n Mr^k \frac{M}{(M + |a_0|)|a_0|^2} \frac{(M + |a_0|)^{n+1-k}}{|a_0|^{n+1-k}} r^{n+1-k} + Mr^{n+1} / |a_0|^2 = \\
&= M' r^{n+1} \frac{M}{|a_0|} \sum_{j=1}^n \left( \frac{M + |a_0|}{|a_0|} \right)^j + Mr^{n+1} / |a_0|^2 \\
&\leq M' r^{n+1} \frac{M}{|a_0|} \left( \left( \frac{M + |a_0|}{|a_0|} \right)^{n+1} - \frac{M + |a_0|}{|a_0|} \right) / \left( \frac{M + |a_0|}{|a_0|} - 1 \right) + Mr^{n+1} / |a_0|^2 = \\
&= M' r^{n+1} \left( \left( \frac{M + |a_0|}{|a_0|} \right)^{n+1} - \frac{M + |a_0|}{|a_0|} \right) + M' r^{n+1} \frac{M + |a_0|}{|a_0|} = M' \rho^{n+1}
\end{aligned}$$

This concludes the proof; we’ll return to clarify the motivation for this proof a bit later.

With the lemma established, we can now conclude, from Hadamard’s formula and the definition of the limsup, that if  $\sum a_n z^n$  has radius of convergence  $\rho_1$ , then for every  $\varepsilon > 0$  there exists some  $M$  such that  $|a_n| \leq M(\frac{1}{\rho_1} + \varepsilon)^n =: Mr^n$ . Now from the estimate for the  $b_n$  provided by the lemma, we conclude that the radius of convergence of  $\sum b_n z^n$  is at least as large as  $1/\rho = |a_0|/[(M + |a_0|)r] = |a_0|/[(M + |a_0|)(\rho_1^{-1} + \varepsilon)]$ . Note that  $M$  depends on  $\varepsilon$ ,

and on all the coefficients  $a_n$ , not only the large ones.

Attempts to simplify the induction step above, even at the price of proving a somewhat weaker estimate, are likely to fail. There is a good reason why this calculation works out with such a precise target landing in the final estimate as it did, and insight comes from *how* this proof was concocted behind the scenery before it was written down.

Basically all convergence results for power series (other than fine results *on* the boundary of the disc of convergence) are based on comparison with geometric series. So we translated the information ‘non-zero radius of convergence into an estimate that is tantamount to comparison with a geometric series: Namely  $|a_n| \leq Mr^n$  means that we are comparing  $\sum a_n z^n$  with the geometric series  $\sum Mz^n$ ; or since we do retain the information about the constant term  $a_0$ , we are comparing with  $a_0 + z \sum Mz^n$ . In a sense, the worst case scenario is when all the coefficients  $a_n$  for  $n \geq 1$  are negative, whereas  $a_0$  is positive; or vice versa. In this case, when calculating the recursion formula for  $b_n$ , namely  $b_n = -\sum_{k=1}^n a_k b_{n-k}/a_0$ , there will be no terms with opposite signs to cancel, and the estimate  $|b_n| \leq \sum |a_k| |b_{n-k}|/|a_0|$  becomes an equality.

In this situation, when  $\sum a_n r^n = a_0 - M \sum_{n=1}^{\infty} r^n = a_0 - Mr/(1-r)$ , we know what the reciprocal series should be: It should be a power series that sums up to the expression  $[a_0 - Mr/(1-r)]^{-1} = \frac{1}{a_0} + \frac{M}{a_0^2} \frac{r}{1-(M+a_0)r/a_0}$ ; this is again a geometric series, and its coefficients are just the quantities that were used as an upper bound for  $|b_n|$  in the lemma.

So in a sense there is one ‘worst case scenario’ in which, when chasing through the recursion formulas for the coefficients of the formal series and putting absolute value signs in, using the triangle inequality, this triangle inequality becomes saturated (i.e., is fulfilled with equality). When this worst case scenario is a precise geometric series (and in all relevant proofs it is), all the handling of coefficients can be calculated explicitly, because the corresponding operations on functions represented by a geometric series can be performed easily; and the new series (in our case reciprocal) represents a worst case scenario for the general situation.

The same proof approach with the same motivation could be used to prove that, plugging in one series  $z = \sum_{n=1}^{\infty} b_n w^n$  (note that we start at  $n = 1$ , not 0) for  $z$  in  $\sum_{n=0}^{\infty} a_n z^n$ , we get again a series that is convergent, at least in a small disc: Indeed, if you plug  $z = b_1 w/(1-Bw)$  into the expression  $a_0 + a_1 z/(1-Az)$ , you get an expression  $a_0 + c_1 w/(1-Cw)$ , where you can easily calculate  $c_1, C$  in terms of  $b_1, B, a_1, A$ . An induction proof for the convergence of the composite of convergent series can be produced according to the same principle as in the case of the reciprocal series, based on this observation the said ‘fractional linear’ functions.

This same method, called ‘majorant method’ applies when using power series to solve ODEs (as we are doing in our 431; e.g., Bessel functions and Bessel equation) or PDEs (as mentioned in our 535 under the label Cauchy-Kovalevskaya theorem) and proving the validity of the formal calculation.