LECTURE NOTES ON NONSTANDARD ANALYSIS
UCLA SUMMER SCHOOL IN LOGIC

ISAAC GOLDBRING

Contents

1. The hyperreals 3
1.1. Basic facts about the ordered real field 3
1.2. The nonstandard extension 4
1.3. Arithmetic in the hyperreals 5
1.4. The structure of $\mathbb{N}^*$ 7
1.5. More practice with transfer 8
1.6. Problems 9
2. Logical formalisms for nonstandard extensions 10
2.1. Approach 1: The compactness theorem 11
2.2. Approach 2: The ultrapower construction 12
2.3. Problems 16
3. Sequences and series 17
3.1. First results about sequences 17
3.2. Cluster points 19
3.3. Series 21
3.4. Problems 22
4. Continuity 23
4.1. First results about continuity 23
4.2. Uniform continuity 25
4.3. Sequences of functions 27
4.4. Problems 30
5. Differentiation 33
5.1. The derivative 33
5.2. Continuous differentiability 35
5.3. Problems 36
6. Riemann Integration 38
6.1. Hyperfinite Riemann sums and integrability 38
6.2. The Peano Existence Theorem 41
6.3. Problems 43
7. Weekend Problem Set #1 44
8. Many-sorted and Higher-Type Structures 47
8.1. Many-sorted structures 47

Date: November 10, 2014.

8.2. Higher-type sorts 48
8.3. Saturation 51
8.4. Useful nonstandard principles 53
8.5. Recap: the nonstandard setting 54
8.6. Problems 54
9. Metric Space Topology 55
9.1. Open and closed sets, compactness, completeness 55
9.2. More about continuity 63
9.3. Compact maps 64
9.4. Problems 65
10. Banach Spaces 67
10.1. Normed spaces 67
10.2. Bounded linear maps 68
10.3. Finite-dimensional spaces and compact linear maps 69
10.4. Problems 71
11. Hilbert Spaces 73
11.1. Inner product spaces 73
11.2. Orthonormal bases and ℓ^2 75
11.3. Orthogonal projections 79
11.4. Hyperfinite-dimensional subspaces 82
11.5. Problems 83
12. Weekend Problem Set #2 85
13. The Spectral Theorem for compact hermitian operators 88
13.1. Problems 93
14. The Bernstein-Robinson Theorem 94
15. Measure Theory 101
15.1. General measure theory 101
15.2. Loeb measure 102
15.3. Product measure 103
15.4. Integration 104
15.5. Conditional expectation 104
15.6. Problems 105
16. Szemerédi Regularity Lemma 106
16.1. Problems 108
References 110

Nonstandard analysis was invented by Abraham Robinson in the 1960s as a way to rescue the naïve use of infinitesimal and infinite elements favored by mathematicians such as Leibniz and Euler before the advent of the rigorous methods introduced by Cauchy and Weierstrauss. Indeed, Robinson realized that the compactness theorem of first-order logic could be used to provide fields that "logically behaved" like the ordered real field while containing "ideal" elements such as infinitesimal and infinite elements.

Since its origins, nonstandard analysis has become a powerful mathematical tool, not only for yielding easier definitions for standard concepts and providing slick proofs of well-known mathematical theorems, but for also providing mathematicians with amazing new tools to prove theorems, e.g. hyperfinite approximation. In addition, by providing useful mathematical heuristics a precise language to be discussed, many mathematical ideas have been elucidated greatly.

In these notes, we try and cover a wide spectrum of applications of nonstandard methods. In the first part of these notes, we explain what a nonstandard extension is and we use it to reprove some basic facts from calculus. We then broaden our nonstandard framework to handle more sophisticated mathematical situations and begin studying metric space topology. We then enter functional analysis by discussing Banach and Hilbert spaces. Here we prove our first serious theorems: the Spectral Theorem for compact hermitian operators and the Bernstein-Robinson Theorem on invariant subspaces; this latter theorem was the first major theorem whose first proof was nonstandard. We then end by briefly discussing Loeb measure and using it to give a slick proof of an important combinatorial result, the Szemerédi Regularity Lemma.

Due to time limitations, there are many beautiful subjects I had to skip. In particular, I had to omit the nonstandard hull construction (although this is briefly introduced in the second weekend problem set) as well as applications of nonstandard analysis to Lie theory (e.g. Hilbert's fifth problem), geometric group theory (e.g. asymptotic cones), and commutative algebra (e.g. bounds in the theory of polynomial rings).

We have borrowed much of our presentation from two main sources: Goldblatt's fantastic book [2] and Davis' concise [1]. Occasionally, I have borrowed some ideas from Henson's [3]. The material on Szemerédi's Regularity Lemma and the Furstenberg Correspondence come from Terence Tao's blog.

I would like to thank Bruno De Mendonca Braga and Jonathan Wolf for pointing out errors in an earlier version of these notes.

1. The hyperreals

1.1. Basic facts about the ordered real field.

The *ordered field of real numbers* is the structure $(\mathbb{R}; +, \cdot, 0, 1, <)$. We recall some basic properties:

- ($\mathbb{Q}$ is dense in $\mathbb{R}$) for every $r \in \mathbb{R}$ and every $\epsilon \in \mathbb{R}^{>0}$, there is $q \in \mathbb{Q}$ such that $|r - q| < \epsilon$;
- (Triangle Inequality) for every $x, y \in \mathbb{R}$, we have $|x + y| \leq |x| + |y|$;
- (Archimedean Property) for every $x, y \in \mathbb{R}^{>0}$, there is $n \in \mathbb{N}$ such that $nx > y$.

Perhaps the most important property of the ordered real field is

Definition 1.1 (Completeness Property)**.** If $A \subseteq \mathbb{R}$ is nonempty and bounded above, then there is a $b \in \mathbb{R}$ such that:

- for all $a \in A$, we have $a \leq b$ (b is an *upper bound for* A);

- if $a \leq c$ for all $a \in A$, then $b \leq c$ (b is the *least* upper bound for A).

Such b is easily seen to be unique and is called the *least upper bound of* A, or the *supremum* of A, and is denote $\sup(A)$.

Exercise 1.2. Show that if A is nonempty and bounded below, then A has a *greatest lower bound*. The greatest lower bound is also called the *infimum of* A and is denoted $\inf(A)$.

1.2. **The nonstandard extension.** In order to start "doing" nonstandard analysis as quickly as possible, we will postpone a formal construction of the nonstandard universe. Instead, we will pose some postulates that a nonstandard universe should possess, assume the existence of such a nonstandard universe, and then begin reasoning in this nonstandard universe. Of course, after we have seen the merits of some nonstandard reasoning, we will return and give a couple of rigorous constructions of nonstandard universes.

We will work in a *nonstandard universe* $\mathbb{R}^*$ that has the following properties:

(NS1) $(\mathbb{R}; +, \cdot, 0, 1, <)$ is an *ordered subfield* of $(\mathbb{R}^*; +, \cdot, 0, 1, <)$.

(NS2) $\mathbb{R}^*$ has a *positive infinitesimal* element, that is, there is $\epsilon \in \mathbb{R}^*$ such that $\epsilon > 0$ but $\epsilon < r$ for every $r \in \mathbb{R}^{>0}$.

(NS3) For every $n \in \mathbb{N}$ and every function $f : \mathbb{R}^n \to \mathbb{R}$, there is a "natural extension" $f : (\mathbb{R}^*)^n \to \mathbb{R}^*$. The natural extensions of the field operations $+, \cdot : \mathbb{R}^2 \to \mathbb{R}$ coincide with the field operations in $\mathbb{R}^*$. Similarly, for every $A \subseteq \mathbb{R}^n$, there is a subset $A^* \subseteq (\mathbb{R}^*)^n$ such that $A^* \cap \mathbb{R}^n = A$.

(NS4) $\mathbb{R}^*$, equipped with the above assignment of extensions of functions and subsets, "behaves logically" like $\mathbb{R}$.

The last item in the above list is, of course, extremely vague and imprecise. We will need to discuss some logic in order to carefully explain what we mean by this. Roughly speaking, any statement that is expressible in first-order logic and mentioning only *standard* numbers is true in $\mathbb{R}$ if and only if it is true in $\mathbb{R}^*$. This is often referred to as the *Transfer Principle*, although, logically speaking, we are just requiring that $\mathbb{R}^*$, in a suitable first-order language, be an *elementary extension* of $\mathbb{R}$. We will explain this in more detail later in these notes.

That being said, until we rigorously explain the logical formalism of nonstandard analysis, we should caution the reader that typical transferrable statements involve quantifiers over *numbers* and **not** *sets* of numbers. For example, the completeness property for $\mathbb{R}$ says that "for all **sets** of numbers A that are nonempty and bounded above, $\sup(A)$ exists." This is an example of a statement that is not transferrable; see Exercise 1.8 below.

Definition 1.3. $\mathbb{R}^*$ is called the *ordered field of hyperreals*.

Remark. If $f : A \to \mathbb{R}$ is a function, where $A \subseteq \mathbb{R}^n$, we would like to also consider its nonstandard extension $f : A^* \to \mathbb{R}^*$. We will take care of this matter shortly.

1.3. **Arithmetic in the hyperreals.** First, let's discuss some immediate consequences of the above postulates. Since $\mathbb{R}^*$ is an ordered field, we can start performing the field operations to our positive infinitesimal ϵ. For example, ϵ has an additive inverse $-\epsilon$, which is then a *negative infinitesimal.* Also, we can consider $\pi \cdot \epsilon$; it is reasonably easy to see that $\pi \cdot \epsilon$ is also a positive infinitesimal. (This will also follow from a more general principle that we will shortly see.)

Since $\epsilon \neq 0$, it has a multiplicative inverse ϵ^{-1}. For a given $r \in \mathbb{R}^{>0}$, since $\epsilon < \frac{1}{r}$, we see that $\epsilon^{-1} > r$. Since r was an arbitrary positive real number, we see that ϵ^{-1} is a *positive infinite element.* And of course, $-\epsilon^{-1}$ is a *negative infinite element.* But now we can continue playing, considering numbers like $\sqrt{2} \cdot \epsilon^{-3}$ and so on...

And besides algebraic manipulations, we also have transcendental matters to consider. Indeed, we have the nonstandard extension of the function $\sin : \mathbb{R}^* \to \mathbb{R}^*$; what is $\sin(\epsilon)$? All in due time... First, let's make precise some of the words we have been thus far freely tossing around.

Definition 1.4.

(1) The set of *finite hyperreals* is
$$\mathbb{R}_{\text{fin}} := \{x \in \mathbb{R}^* \mid |x| \leq n \text{ for some } n \in \mathbb{N}\}.$$
(2) The set of *infinite hyperreals* is $\mathbb{R}_{\text{inf}} := \mathbb{R}^* \setminus \mathbb{R}_{\text{fin}}$.
(3) The set of *infinitesimal hyperreals* is
$$\mu := \{x \in \mathbb{R}^* \mid |x| \leq \frac{1}{n} \text{ for all } n \in \mathbb{N}^{>0}\}.$$

The notation μ comes from the more general notion of *monad*, which we will encounter later in the notes.

Observe that $\mu \subseteq \mathbb{R}_{\text{fin}}$, $\mathbb{R} \subseteq \mathbb{R}_{\text{fin}}$, and $\mu \cap \mathbb{R} = \{0\}$. Also note that if $\delta \in \mu \setminus \{0\}$, then $\delta^{-1} \notin \mathbb{R}_{\text{fin}}$.

Lemma 1.5.

(1) $\mathbb{R}_{\text{fin}}$ *is a* subring *of* $\mathbb{R}^*$*: for all* $x, y \in \mathbb{R}_{\text{fin}}$, $x \pm y$, $x \cdot y \in \mathbb{R}_{\text{fin}}$.
(2) μ *is an* ideal *of* $\mathbb{R}_{\text{fin}}$*:* μ *is a subring of* $\mathbb{R}^*$ *and for all* $x \in \mathbb{R}_{\text{fin}}$ *and* $y \in \mu$*, we have* $xy \in \mu$.

Proof. (1) Fix $x, y \in \mathbb{R}_{\text{fin}}$. Choose $r, s \in \mathbb{R}^{>0}$ such that $|x| \leq r$ and $|y| \leq s$. Then $|x \pm y| \leq r + s$ and $|xy| \leq rs$, whence $x \pm y$, $xy \in \mathbb{R}_{\text{fin}}$.

(2) Suppose $x, y \in \mu$. We need to show that $x \pm y \in \mu$. Fix $r \in \mathbb{R}^{>0}$; we need $|x \pm y| \leq r$. Well, since $x, y \in \mu$, we have that $|x|, |y| \leq \frac{r}{2}$. Then, $|x \pm y| \leq |x| + |y| \leq \frac{r}{2} + \frac{r}{2} = r$.

Now suppose $x \in \mathbb{R}_{\text{fin}}$ and $y \in \mu$. We need $xy \in \mu$. Fix $r \in \mathbb{R}^{>0}$; we need $|xy| \leq r$. Choose $s \in \mathbb{R}^{>0}$ such that $|x| \leq s$. Since $y \in \mu$, we have $|y| \leq \frac{r}{s}$. Thus, $|xy| = |x||y| \leq s \cdot \frac{r}{s} = r$. $\square$

A natural question now arises: What is the quotient ring $\mathbb{R}_{\text{fin}}/\mu$? The answer will arrive shortly.

Definition 1.6. For $x, y \in \mathbb{R}^*$, we say x and y are *infinitely close*, written, $x \approx y$, if $x - y \in \mu$.

Exercise 1.7.

(1) $\approx$ is an equivalence relation on $\mathbb{R}^*$, namely, for all $x, y, z \in \mathbb{R}^*$:
  - $x \approx x$;
  - $x \approx y$ implies $y \approx x$;
  - $x \approx y$ and $y \approx z$ implies $x \approx z$.

(2) $\approx$ is a congruence relation on $\mathbb{R}_{\text{fin}}$, namely, $\approx$ is an equivalence relation on $\mathbb{R}_{\text{fin}}$ and, for all $x, y, u, v \in \mathbb{R}_{\text{fin}}$, if $x \approx u$ and $y \approx v$, then $x \pm y \approx u \pm v$ and $xy \approx uv$.

Exercise 1.8. Show that $\mathbb{R}$ is a nonempty set that is bounded above in $\mathbb{R}^*$ but that $\sup(\mathbb{R})$ does not exist in $\mathbb{R}^*$. Thus, the completeness property is not true for $\mathbb{R}^*$.

The next theorem is of extreme importance.

Theorem 1.9 (Existence of Standard Parts)**.** *If $r \in \mathbb{R}_{\text{fin}}$, then there is a unique $s \in \mathbb{R}$ such that $r \approx s$. We call s the* standard part of *r and write* $\text{st}(r) = s$.

Proof. Uniqueness is immediate: if $r \approx s_1$ and $r \approx s_2$, with $s_1, s_2 \in \mathbb{R}$, then $s_1 \approx s_2$, so $s_1 - s_2 \in \mu \cap \mathbb{R} = \{0\}$, so $s_1 = s_2$. We now show existence. Without loss of generality, we can assume $r > 0$. (Why?) We then set $A = \{x \in \mathbb{R} \mid x < r\}$. Since $r \in \mathbb{R}_{\text{fin}}$, A is bounded above. Also, $0 \in A$, so A is nonempty. Thus, by the Completeness Property, $\sup(A)$ exists. Set $s := \sup(A)$. We claim that this is the desired s. Fix $\delta \in \mathbb{R}^{>0}$. Since s is an upper bound for A, $s + \delta \notin A$, so $r \leq s + \delta$. If $r \leq s - \delta$, then $s - \delta$ would be an upper bound for A, contradicting the fact that s was the least upper bound for A. Consequently, $r \geq s - \delta$. It follows that $|r - s| \leq \delta$. Since $\delta \in \mathbb{R}^{>0}$ is arbitrary, it follows that $r - s \in \mu$. $\square$

Remark. In proving that standard parts exist, we used the Completeness Property for $\mathbb{R}$. Later, we will show that the Completeness Property is a consequence of the existence of standard parts. Thus, an equivalent way to state the completeness property for $\mathbb{R}$ is that standard parts exist.

Remark. Later, when we start studying metric space topology from the nonstandard perspective, we will call an element *nearstandard* if it is within an infinitesimal distance from a standard element. Some metric spaces also have a natural notion of *finite*, e.g. $\mathbb{R}$, and, more generally, normed vector spaces. The equivalence between finite and nearstandard asserted in the previous theorem will also hold for finite-dimensional normed spaces. In general, nearstandard points are always finite, but in infinite-dimensional settings, this inclusion is often strict. This topic will be discussed later in these notes.

Exercise 1.10. Let $x, y \in \mathbb{R}_{\text{fin}}$.

(1) $x \approx y$ if and only if $\mathrm{st}(x) = \mathrm{st}(y)$.
(2) If $x \leq y$, then $\mathrm{st}(x) \leq \mathrm{st}(y)$. The converse of this statement is false; give an example.
(3) If $x \in \mathbb{R}$, then $\mathrm{st}(x) = x$.

Theorem 1.11. $\mathrm{st} : \mathbb{R}_{\mathrm{fin}} \to \mathbb{R}$ *is a surjective* ring homomorphism*: for all* $x, y \in \mathbb{R}_{\mathrm{fin}}$*,* $\mathrm{st}(x+y) = \mathrm{st}(x) + \mathrm{st}(y)$ *and* $\mathrm{st}(xy) = \mathrm{st}(x)\,\mathrm{st}(y)$*.*

Proof. This follows immediately from Exercises 1.7(2) and 1.10(3). □

Corollary 1.12. $\mathbb{R}_{\mathrm{fin}}/\mu \cong \mathbb{R}$*.*

Proof. The kernel of st is precisely μ; now use the First Isomorphism Theorem for rings. □

Corollary 1.13. μ *is a* maximal *ideal of* $\mathbb{R}_{\mathrm{fin}}$*.*

Proof. This follows from the fact that $\mathbb{R}_{\mathrm{fin}}/\mu$ is a field. □

Exercise 1.14. Give a direct proof of the last corollary, that is, show directly that μ is a maximal ideal of $\mathbb{R}_{\mathrm{fin}}$.

1.4. **The structure of $\mathbb{N}^*$.** In this subsection, let's take a brief look at the picture of $\mathbb{N}^*$. First, let's establish that $\mathbb{N}^* \setminus \mathbb{N} \neq \emptyset$. To see this, let $y \in \mathbb{R}^*$ be positive infinite. Since the statement "for all $x \in \mathbb{R}$, if $x > 0$, then there is $n \in \mathbb{N}$ such that $x \leq n$" is true in $\mathbb{R}$, the statement "for all $x \in \mathbb{R}^*$, if $x > 0$, then there is $n \in \mathbb{N}^*$ such that $x \leq n$" is true in $\mathbb{R}^*$ by the transfer principle. Thus, there is $N \in \mathbb{N}^*$ such that $y \leq N$. However, if $N \in \mathbb{N}$, then y is finite, a contradiction. Thus, $N \in \mathbb{N}^* \setminus \mathbb{N}$. Also note that the same argument implies that N is positive infinite.

The last sentence of the previous paragraph holds for all $N \in \mathbb{N}^* \setminus \mathbb{N}$: if $N \in \mathbb{N}^* \setminus \mathbb{N}$, then N is positive infinite. Indeed, the statement "for all $n \in \mathbb{N}$, $n \geq 0$" is true in $\mathbb{R}$, so the statement "for all $n \in \mathbb{N}^*$, $n \geq 0$" is true in $\mathbb{R}^*$, whence $N \geq 0$. Also, if $N \in \mathbb{R}_{\mathrm{fin}}$, then there is $n \in \mathbb{N}$ such that $n \leq N \leq n+1$. However, the statement "for all $m \in \mathbb{N}$, if $n \leq m \leq n+1$, then $m = n$ or $m = n+1$" is true in $\mathbb{R}$; applying the transfer principle to this statement, we have $N = n$ or $N = n+1$, whence $N \in \mathbb{N}$, a contradiction. Thus, $N \in \mathbb{R}_{\mathrm{inf}}$.

Now that we know that all *nonstandard natural numbers* are positive infinite, let's ask the question: "How many nonstandard natural numbers are there?" To examine this question, let's first establish some notation and terminology. For $N \in \mathbb{N}^*$, set $\gamma(N) := \{N \pm m \mid m \in \mathbb{N}\}$, the *galaxy* or *archimedean class* of N. Clearly, $N \in \mathbb{N}$ if and only if $\gamma(N) = \mathbb{Z}$; we call this the *finite galaxy*, while all other galaxies will be referred to as *infinite galaxies*.

Lemma 1.15. *If* $N \in \mathbb{N}^* \setminus \mathbb{N}$*, then* $\gamma(N) \subseteq \mathbb{N}^*$*.*

Proof. By transfer, $N+1 \in \mathbb{N}^*$; by induction, this shows that $N+m \in \mathbb{N}^*$ for all $m \in \mathbb{N}$. We now show, also inductively, that $N - m \in \mathbb{N}^*$ for all $m \in \mathbb{N}$.

Suppose that the result is true for a given m. Notice that $N - m \neq 0$, else $N = m \in \mathbb{N}$. Applying transfer to the statement "for all $n \in \mathbb{N}$, if $n \neq 0$, then $n - 1 \in \mathbb{N}$," we see that $(N - m) - 1 = N - (m + 1) \in \mathbb{N}^*$. $\square$

Since we know we have at least one nonstandard natural number, we now know that we have an entire galaxy of them. Notice that a galaxy looks just like a copy of $\mathbb{Z}$ and that $\gamma(M) = \gamma(N)$ if and only if $|M - N| \in \mathbb{N}$.

Observe that if $\gamma(M) = \gamma(M')$ and $\gamma(N) = \gamma(N')$ and $\gamma(M) \neq \gamma(N)$, then $M < N$ if and only if $M' < N'$. Consequently, we can define an ordering on galaxies: if $\gamma(M) \neq \gamma(N)$, then we say $\gamma(M) < \gamma(N)$ if and only if $M < N$. When $\gamma(M) < \gamma(N)$, we think of M as being *infinitely less than N*.

What can be said about the ordering of the set of galaxies? In particular, are there more than just two galaxies?

Lemma 1.16. *The set of* infinite *galaxies is densely ordered without endpoints, meaning:*

1. *there is no largest infinite galaxy, that is, for every $M \in \mathbb{N}^* \setminus \mathbb{N}$, there is $N \in \mathbb{N}^* \setminus \mathbb{N}$ such that $\gamma(M) < \gamma(N)$;*
2. *there is no smallest infinite galaxy, that is, for every $M \in \mathbb{N}^* \setminus \mathbb{N}$, there is $N \in \mathbb{N}^* \setminus \mathbb{N}$ such that $\gamma(N) < \gamma(M)$;*
3. *between any two infinite galaxies, there is a third (infinite) galaxy, that is, for every $M_1, M_2 \in \mathbb{N}^* \setminus \mathbb{N}$ such that $\gamma(M_1) < \gamma(M_2)$, there is $N \in \mathbb{N}^* \setminus N$ such that $\gamma(M_1) < \gamma(N) < \gamma(M_2)$.*

Proof. (1) Given $M \in \mathbb{N}^* \setminus \mathbb{N}$, we claim that $\gamma(M) < \gamma(2M)$. Otherwise, $2M = M + m$ for some $m \in \mathbb{N}$, whence $M = m$, a contradiction.

(2) Since $\gamma(M) = \gamma(M-1)$, we may as well suppose that M is even. Then $\gamma(\frac{M}{2}) < \gamma(M)$ from the proof of (1); it remains to note that $\frac{M}{2} \in \mathbb{N}^* \setminus \mathbb{N}$.

(3) Again, we may as well assume that M_1 and M_2 are both even. In this case, arguing as before, one can see that $\gamma(M_1) < \gamma(\frac{M_1+M_2}{2}) < \gamma(M_2)$. $\square$

Under suitable richness assumptions on the nonstandard extension (to be discussed later), one can go even further: if $(N_\alpha)_{\alpha<\kappa}$ is a descending sequence of nonstandard natural numbers, then there is $N \in \mathbb{N}^* \setminus \mathbb{N}$ such that $N < N_\alpha$ for all $\alpha < \kappa$.

1.5. **More practice with transfer.** In order to get some practice with the Transfer Principle, we will prove the assertion made in Remark 1.3. More precisely:

Theorem 1.17. *The statement "every finite element of $\mathbb{R}^*$ has a standard part" implies the Completeness Property of the ordered real field.*

Proof. Suppose that $A \subseteq \mathbb{R}$ is nonempty and bounded above. We must show that $\sup(A)$ exists. Let $b \in \mathbb{R}$ be an upper bound for A. Let's define a function $f : \mathbb{R} \to \mathbb{R}$ as follows: If $r \in \mathbb{R} \setminus \mathbb{N}$, set $f(r) = 0$. Otherwise, set $f(n) =$ the least $k \in \mathbb{Z}$ such that $\frac{k}{n}$ is an upper bound for A; such a k exists by the Archimedean property. Observe that, if $m, n \in \mathbb{N}$ and $n \leq m$, then $f(n) \leq f(m)$.

We now consider the nonstandard extension of the function $f : \mathbb{R}^* \to \mathbb{R}^*$. Fix $N \in \mathbb{N}^* \setminus \mathbb{N}$. The key idea is to understand $f(N)$. By the transfer principle, we know that $f(N)$ is the least element of $\mathbb{Z}^*$ such that $\frac{f(N)}{N}$ is an upper bound for A^*. Also, by the transfer principle applied to the last observation of the previous paragraph, if $m \in \mathbb{N}$, then $f(m) \leq f(N)$.

Claim 1: $\frac{f(N)}{N} \in \mathbb{R}_{\text{fin}}$: Suppose this is not the case, so, $\frac{f(N)}{N} \in \mathbb{R}_{\text{inf}}$. Since $f(1) \leq f(N)$, we must have that $f(N)$ is a *positive* infinite element. Since the statement "for all $n \in \mathbb{N}$, there is $a \in A$ such that $\frac{f(n)-1}{n} < a$" is true in $\mathbb{R}$, by the Transfer Principle, the statement for all $n \in \mathbb{N}^*$, there is $a \in A^*$ such that $\frac{f(n)-1}{n} < a$" is true in $\mathbb{R}^*$. Thus, we may fix $a_0 \in A^*$ such that $\frac{f(N)-1}{N} < a_0$. Since $\frac{f(N)-1}{N}$ is a positive infinite element (as it differs from the positive infinite element $\frac{f(N)}{N}$ by the infinitesimal amount $\frac{1}{N}$), we have that a_0 is also a positive infinite element of $\mathbb{R}^*$. However, the statement "for all $a \in A$, we have $a \leq b$" is true in $\mathbb{R}$, whence the statement "for all $a \in A^*$, we have $a \leq b$" is true in $\mathbb{R}^*$. Consequently, $a_0 \leq b$, contradicting the fact that a is a positive infinite element of $\mathbb{R}^*$.

By Claim 1 and the assumption of the theorem, $\text{st}(\frac{f(N)}{N})$ exists. Set $r := \text{st}(\frac{f(N)}{N})$.

Claim 2: r is an upper bound for A: To see this, fix $a \in A$. Then, by the transfer principle, $r \leq \frac{f(N)}{N}$, whence $a = \text{st}(a) \leq \text{st}(\frac{f(N)}{N}) = r$.

Claim 3: $r = \sup(A)$: To see this, fix $\delta \in \mathbb{R}^{>0}$. We must find $a \in A$ such that $r - \delta < a$. As we showed in the proof of Claim 1, there is $a \in A^*$ such that $\frac{f(N)-1}{N} < a$. Since $\frac{1}{N} < \delta$, we have $r - \delta < \frac{f(N)-1}{N}$, whence $r - \delta < a$. In other words, the statement "there is $a \in A^*$ such that $r - \delta < a$" is true in $\mathbb{R}^*$. By the transfer principle, the statement "there is $a \in A$ such that $r - \delta < a$" is true in $\mathbb{R}$, which is precisely what we needed. □

1.6. **Problems.**

Problem 1.1. Let $x, x', y, y' \in \mathbb{R}_{\text{fin}}$.

(1) Show that $x \approx y$ if and only if $\text{st}(x) = \text{st}(y)$.
(2) Show that if $x \in \mathbb{R}$, then $x = \text{st}(x)$.
(3) Show that $x \leq y$ implies $\text{st}(x) \leq \text{st}(y)$. Show that the converse is false.
(4) Show that if $\text{st}(x) < \text{st}(y)$, then $x < y$. In fact, show that if $\text{st}(x) < \text{st}(y)$, then there is $r \in \mathbb{R}$ such that $x < r < y$.
(5) Suppose that $x \approx x'$ and $y \approx y'$. Show that:
   (a) $x \pm y \approx x' \pm y'$;

(b) $x \cdot y \approx x' \cdot y'$;
(c) $\frac{x}{y} \approx \frac{x'}{y'}$ if $y \not\approx 0$.

Show that (c) can fail if $y, y' \in \mu \setminus \{0\}$.

Problem 1.2. Suppose $x, y \in \mathbb{R}^*$ and $x \approx y$. Show that if $b \in \mathbb{R}_{\text{fin}}$, then $bx \approx by$. Show that this can fail if $b \notin \mathbb{R}_{\text{fin}}$.

Problem 1.3.
1. Show that $\mathbb{R}^*$ is not complete by finding a nonempty subset of $\mathbb{R}^*$ which is bounded above that does not have a supremum.
2. Show that if $A \subseteq \mathbb{R}$ is unbounded, then A has no least upper bound when considered as a subset of $\mathbb{R}^*$. (This may even be how you solved part (a).)

Problem 1.4. Let F be an ordered field. F is said to be *archimedean* if for any $x, y \in F$ with $x, y > 0$, there is $n \in \mathbb{N}$ such that $y < nx$. Show that $\mathbb{R}^*$ is not archimedean. (It is a fact that archimedean ordered fields are complete, so this problem strengthens the result of the previous problem.)

Problem 1.5. Construct a sequence of subsets (A_n) of $\mathbb{R}$ such that

$$(\bigcup_{n=1}^{\infty} A_n)^* \neq \bigcup_{n=1}^{\infty} A_n^*.$$

Problem 1.6. If F is a field and V is an F-vector space, let $\dim_F(V)$ denote the dimension of V as an F-vector space.
1. Observe that $\mathbb{R}^*$ is an $\mathbb{R}^*$-vector space. (More generally, any field F is naturally an F-vector space.) What is $\dim_{\mathbb{R}^*}(\mathbb{R}^*)$?
2. Observe that $\mathbb{R}^*$ is also a vector space over $\mathbb{R}$. Show that $\dim_{\mathbb{R}}(\mathbb{R}^*) = \infty$. (Hint: Let $x \in \mathbb{R}^* \setminus \mathbb{R}$. Show that $\{1, x, x^2, \ldots, x^n, \ldots\}$ is an $\mathbb{R}$-linearly independent set.)
3. Show that $\mathbb{R}_{\text{fin}}$ is an $\mathbb{R}$-subspace of $\mathbb{R}^*$. What is $\dim_{\mathbb{R}}(\mathbb{R}_{\text{fin}})$?
4. Show that μ is an $\mathbb{R}$-subspace of $\mathbb{R}^*$. What is $\dim_{\mathbb{R}}(\mu)$?
5. What is $\dim_{\mathbb{R}}(\mathbb{R}^*/\mathbb{R}_{\text{fin}})$?
6. What is $\dim_{\mathbb{R}}(\mathbb{R}_{\text{fin}}/\mu)$?

Problem 1.7. Show that $\text{card}(\mathbb{N}^*) \geq 2^{\aleph_0}$; here $\text{card}(A)$ denotes the cardinality of the set A. (Hint: First show that $\text{card}(\mathbb{Q}^*) \geq 2^{\aleph_0}$).

Problem 1.8. Give a direct proof that μ is a maximal ideal of $\mathbb{R}_{\text{fin}}$, that is, show that if I is an ideal of $\mathbb{R}_{\text{fin}}$ such that $\mu \subseteq I$, then $I = \mu$ or $I = \mathbb{R}_{\text{fin}}$.

2. Logical formalisms for nonstandard extensions

At this point, you might be wondering one of two things:
1. What else can I do with these wonderful postulates for nonstandard extensions? or
2. Does such a nonstandard extension exist or was everything done in Section 1 all magical nonsense?

If you asked the former question, you can safely skip this section and discover the wonders of the nonstandard calculus to come in the following sections. (But please, at some point, return and read this section!) If you asked the latter question, we will ease your trepidations by offering not one, but two, different logical formalisms for nonstandard extensions. The first formalism will rely heavily on the Compactness Theorem from first-order logic, but, modulo that prerequisite, this route is the quickest way to obtain nonstandard extensions. The second formalism is the Ultraproduct Approach, which is the most algebraic and "mainstream" way to explain nonstandard methods to "ordinary" mathematicians. Of course, at some point, some logic must be introduced in the form of Łos' (pronounced "Wash's") theorem, which will be discussed as well.

2.1. **Approach 1: The compactness theorem.** In this section, some familiarity with first-order logic is assumed. We let $\mathcal{L}$ denote the first-order language consisting of the following symbols:

- for every $r \in \mathbb{R}$, we have a constant symbol c_r;
- for every $n > 0$ and every $f : \mathbb{R}^n \to \mathbb{R}$, we have a n-ary function symbol F_f;
- for every $n > 0$ and every $A \subseteq \mathbb{R}^n$, we have a n-ary relation symbol P_A.

We let $\mathfrak{R}$ be the $\mathcal{L}$-structure whose universe is $\mathbb{R}$ and whose symbols are interpreted in the natural way, namely $(c_r)^{\mathfrak{R}} = r$, $(F_f)^{\mathfrak{R}} = f$, and $(P_A)^{\mathfrak{R}} = A$. Let $\Gamma = \mathrm{Th}(\mathfrak{R}) \cup \{c_0 < v < c_r \mid r \in \mathbb{R}^{>0}\}$, a set of $\mathcal{L}$-formulae. Then Γ is finitely satisfiable: for any $r_1, \ldots, r_n \in \mathbb{R}^{>0}$, choose $r \in \mathbb{R}^{>0}$ such that $r < r_i$ for $i = 1, \ldots, n$. Then $\mathfrak{R} \models \mathrm{Th}(\mathfrak{R}) \cup \{c_0 < v < c_{r_i} \mid i = 1, \ldots, n\}[\![r]\!]$. By the Compactness Theorem, Γ is satisfiable, say by some $\mathcal{L}$-structure $\mathfrak{A}$ and $a \in A$. Intuitively, a will be our desired positive infinitesimal. But first, we need $\mathfrak{R}$ to be a substructure of $\mathfrak{A}$; some "abstract nonsense" can help take care of this.

Exercise 2.1.

(1) The function $h : \mathfrak{R} \to \mathfrak{A}$ given by $h(r) = (c_r)^{\mathfrak{A}}$ is an injective homomorphism of $\mathcal{L}$-structures.
(2) Use (1) to find some $\mathcal{L}$-structure $\mathfrak{A}'$ isomorphic to $\mathfrak{A}$ such that $\mathfrak{R}$ is a substructure of $\mathfrak{A}'$.

By the result of the previous exercise, we may suppose that $\mathfrak{R}$ is a substructure of $\mathfrak{A}$. In this case, we denote $\mathfrak{A}$ by $\mathfrak{R}^*$ and denote the universe of $\mathfrak{R}^*$ by $\mathbb{R}^*$. For $f : \mathbb{R}^n \to \mathbb{R}$, we let $f : (\mathbb{R}^*)^n \to \mathbb{R}^*$ denote $(F_f)^{\mathfrak{R}^*}$. For $A \subseteq \mathbb{R}^n$, we set $A^* := (P_A)^{\mathfrak{R}^*}$. These are the extensions that axiom (NS3) postulates. Since being an ordered field is part of $\mathrm{Th}(\mathfrak{R})$, we have that $\mathbb{R}^*$ is an ordered field and, since $\mathfrak{R}$ is a substructure of $\mathfrak{R}^*$, we have that $\mathbb{R}$ is an ordered subfield of $\mathbb{R}^*$, verifying postulate (NS1). Let $\epsilon \in \mathbb{R}^*$ be such that, for every $r \in \mathbb{R}^{>0}$, $\mathfrak{R}^* \models c_0 < v < c_r[\![\epsilon]\!]$. Then ϵ is a positive infinitesimal, verifying postulate (NS2). Finally, the fact that $\mathfrak{R}^* \models \mathrm{Th}(\mathfrak{R})$ is the rigorous, precise meaning of postulate (NS4). We have thus proven:

Theorem 2.2. *There is a nonstandard universe, namely* $\mathfrak{R}^*$.

Exercise 2.3. Suppose that $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$ is a function.

(1) Let $f_1 : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary extension of f to all of $\mathbb{R}^n$. Define $f^* := f_1^* | A^*$. Show that f^* is independent of the choice of extension f_1. (This provides a way to define the nonstandard extensions of partial functions.)

(2) Set $\Gamma_f := \{(x, y) \in \mathbb{R}^{n+1} \mid f(x) = y\}$. Show that Γ_f^* is the graph of a function $g : A^* \to \mathbb{R}^*$. Show also that $g = f^*$.

The following observation is very useful.

Proposition 2.4. *Suppose that* $\varphi(x_1, \ldots, x_m)$ *is an* $\mathcal{L}$*-formula. Set* $S := \{\vec{r} \in \mathbb{R}^m \mid \mathfrak{R} \models \varphi[\![\vec{r}]\!]\}$. *Then* $S^* = \{\vec{r} \in (\mathbb{R}^*)^m \mid \mathfrak{R}^* \models \varphi[\![\vec{r}]\!]\}$.

Proof. Just observe that $\forall \vec{v}(P_S \vec{v} \leftrightarrow \varphi(\vec{v}))$ belongs to $\mathrm{Th}(\mathfrak{R})$. □

Corollary 2.5. $\mathbb{N}$ *is not a definable set (even with parameters) in* $\mathfrak{R}^*$.

Proof. Suppose, towards a contradiction, that there is an $\mathcal{L}$-formula $\varphi(x, \vec{y})$ and $\vec{r} \in \mathbb{R}^*$ such that $\mathbb{N} = \{a \in \mathbb{R}^* \mid \mathfrak{R}^* \models \varphi[\![a, \vec{r}]\!]\}$. Write down an $\mathcal{L}$-sentence σ which says that for all $\vec{y}$, if $\varphi(x, \vec{y})$ defines a nonempty set of natural numbers that is bounded above, then $\varphi(x, \vec{y})$ has a maximum. (Remember you have a symbol $P_{\mathbb{N}}$ for the set of natural numbers.) Since $\mathfrak{R} \models \sigma$, we have $\mathfrak{R}^* \models \sigma$. Now $\mathbb{N}$ is bounded above in $\mathbb{R}^*$ (by an infinite element). Thus, $\mathbb{N}$ should have a maximum in $\mathbb{R}^*$, which is clearly ridiculous. □

Since $\mathbb{N}^*$ is a definable set in $\mathfrak{R}^*$ (defined by $P_{\mathbb{N}}$), we obtain the following

Corollary 2.6. $\mathbb{N}^* \setminus \mathbb{N}$ *is not definable in* $\mathfrak{R}^*$.

In modern nonstandard analysis parlance, the previous two results would be phrased as "$\mathbb{N}$ and $\mathbb{N}^* \setminus \mathbb{N}$" are not *internal* sets. We will discuss internal sets later in these notes.

2.2. **Approach 2: The ultrapower construction.** In this approach to nonstandard analysis, one gives an "explicit" construction of the nonstandard universe in a manner very similar to the explicit construction of the real numbers from the rational numbers. Recall that a real number can be viewed as a sequence of rational numbers which we view as better and better approximations to the real number. Similarly, an element of $\mathbb{R}^*$ should be viewed as a sequence of real numbers. For example, the sequence $(1, 2, 3, \ldots)$ should represent some *infinite* element of $\mathbb{R}^*$.

However, many different sequences of rational numbers represent the same real number. Thus, a real number is an equivalence class of sequences of rational numbers (q_n), where (q_n) and (q'_n) are equivalent if they "represent the same real number," or, more formally, if $\lim_{n\to\infty}(q_n - q'_n) = 0$. We run into the same issue here: many sequences of real numbers should represent the same hyperreal number. For instance, it should hopefully be clear that the sequence $(1, 2, 3, \ldots, n, n+1, \ldots)$ and $(\pi, e, -72, 4, 5, 6, \ldots, n, n+1, \ldots)$ should represent the same (infinite) hyperreal number as they only differ in a finite number of coordinates.

More generally, we would like to say that two sequences of real numbers represent the same hyperreal number if they agree on "most" coordinates. But what is a good notion of "most" coordinates? A first guess might be that "most" means all but finitely many; it turns out that this guess is insufficient for our purposes. Instead, we will need a slightly more general notion of when two sequences agree on a large number of coordinates; this brings in the notion of a filter.

Definition 2.7. A (proper) *filter* on $\mathbb{N}$ is a set $\mathcal{F}$ of subsets of $\mathbb{N}$ (that is, $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$) such that:

- $\emptyset \notin \mathcal{F}$, $\mathbb{N} \in \mathcal{F}$;
- if $A, B \in \mathcal{F}$, then $A \cap B \in F$;
- if $A \in \mathcal{F}$ and $A \subseteq B$, then $B \in \mathcal{F}$.

We think of elements of $\mathcal{F}$ as "big" sets (because that's what filters do, they catch the big objects). The first and third axioms are (hopefully) intuitive properties of big sets. Perhaps the second axiom is not as intuitive, but if one thinks of the complement of a big set as a "small" set, then the second axiom asserts that the union of two small sets is small (which is hopefully more intuitive).

Exercise 2.8. Set $\mathcal{F} := \{A \subseteq \mathbb{N} \mid \mathbb{N} \setminus A \text{ is finite}\}$. Prove that $\mathcal{F}$ is a filter on $\mathbb{N}$, called the *Frechet* or *cofinite* filter on $\mathbb{N}$.

Exercise 2.9. Suppose that $\mathcal{D}$ is a set of subsets of $\mathbb{N}$ with the *finite intersection property*, namely, whenever $D_1, \ldots, D_n \in \mathcal{D}$, we have $D_1 \cap \cdots \cap D_n \neq \emptyset$. Set

$$\langle \mathcal{D} \rangle := \{E \subseteq \mathbb{N} \mid D_1 \cap \cdots D_n \subseteq E \text{ for some } D_1, \ldots, D_n \in \mathcal{D}\}.$$

Show that $\langle \mathcal{D} \rangle$ is the smallest filter on $\mathbb{N}$ containing $\mathcal{D}$, called the *filter generated by* $\mathcal{D}$.

If $\mathcal{F}$ is a filter on $\mathbb{N}$, then a subset of $\mathbb{N}$ cannot be simultaneously big and small (that is, both it and it's complement belong to $\mathcal{F}$), but there is no requirement that it be one of the two. It will be desirable (for reasons that will become clear in a second) to add this as an additional property:

Definition 2.10. If $\mathcal{F}$ is a filter on $\mathbb{N}$, then $\mathcal{F}$ is an *ultrafilter* if, for any $A \subseteq \mathbb{N}$, either $A \in \mathcal{F}$ or $\mathbb{N} \setminus A \in \mathcal{F}$ (but not both!).

Ultrafilters are usually denoted by $\mathcal{U}$. Observe that the Frechet filter on $\mathbb{N}$ is not an ultrafilter since there are sets $A \subseteq \mathbb{N}$ such that A and $\mathbb{N} \setminus A$

are both infinite (e.g. the even numbers). So what is an example of an ultrafilter on $\mathbb{N}$?

Definition 2.11. Given $m \in \mathbb{N}$, set $\mathcal{F}_m := \{A \subseteq \mathbb{N} \mid m \in A\}$.

Exercise 2.12. For $m \in \mathbb{N}$, prove that $\mathcal{F}_m$ is an ultrafilter on $\mathbb{N}$, called the *principal ultrafilter generated by* m.

We say that an ultrafilter $\mathcal{U}$ on $\mathbb{N}$ is *principal* if $\mathcal{U} = \mathcal{F}_m$ for some $m \in \mathbb{N}$. Although principal ultrafilters settle the question of the existence of ultrafilters, they will turn out to be useless for our purposes, as we will see in a few moments.

Exercise 2.13. Prove that an ultrafilter $\mathcal{U}$ on $\mathbb{N}$ is principal if and only if there is a *finite* set $A \subseteq \mathbb{N}$ such that $A \in \mathcal{U}$.

Exercise 2.14. Suppose that $\mathcal{U}$ is an ultrafilter on $\mathbb{N}$ and $A_1, \ldots, A_n$ are pairwise disjoint subsets of $\mathbb{N}$ such that $A_1 \cup \cdots \cup A_n \in \mathcal{U}$. Prove that there is a unique $i \in \{1, \ldots, n\}$ such that $A_i \in \mathcal{U}$.

We are now ready to explain the ultrapower construction. Fix an ultrafilter $\mathcal{U}$ on $\mathbb{N}$. If (a_n) and (b_n) are infinite sequences of real numbers, we say that (a_n) and (b_n) are *equal modulo* $\mathcal{U}$, written $(a_n) \sim_{\mathcal{U}} (b_n)$, if $\{n \in \mathbb{N} | \mid a_n = b_n\} \in \mathcal{U}$. (This is the precise meaning of when two sequences agree on "most" coordinates.)

Exercise 2.15. Show that $\sim_{\mathcal{U}}$ is an equivalence relation on the set of infinite sequences of real numbers. (The "ultra" assumption is not used in this exercsie.)

For an infinite sequence (a_n), we write $[(a_n)]_{\mathcal{U}}$, or simply $[(a_n)]$, for the equivalence class of (a_n) with respect to $\sim_{\mathcal{U}}$. We let $\mathbb{R}^{\mathcal{U}}$ denote the set of $\sim_{\mathcal{U}}$-equivalence classes.

We want to turn $\mathbb{R}^{\mathcal{U}}$ into an ordered field. The natural guess for the field operations are:

- $[(a_n)] + [(b_n)] := [(a_n + b_n)]$
- $[(a_n)] \cdot [(b_n)] := [(a_n b_n)]$.

Of course, since we are working with equivalence classes, we need to make sure that the above operations are well-defined. For example, if $(a_n) \sim_{\mathcal{U}} (a'_n)$ and $(b_n) \sim_{\mathcal{U}} (b'_n)$, we need to know that $(a_n + b_n) \sim_{\mathcal{U}} (a'_n + b'_n)$ (and similarly for multiplication). However, if we set $A := \{n \in \mathbb{N} \mid a_n = a'_n\}$ and $B := \{n \in \mathbb{N} \mid b_n = b'_n\}$, then for $n \in A \cap B$, $a_n + b_n = a'_n + b'_n$; since $A \cap B \in \mathcal{U}$ and $A \cap B \subseteq \{n \in \mathbb{N} \mid a_n + b_n = a'_n + b'_n\}$, this shows that $(a_n + b_n) \sim_{\mathcal{U}} (a'_n + b'_n)$.

Next we define an order on $\mathbb{R}^{\mathcal{U}}$ by declaring $[(a_n)] < [(b_n)]$ if and only if $\{n \in \mathbb{N} \mid a_n < b_n\} \in \mathcal{U}$. One needs to verify that this is in fact a linear order. For example, to check the "linearity" axioms, fix $[(a_n)], [(b_n)] \in \mathbb{R}^{\mathcal{U}}$. Set $A := \{n \in \mathbb{N} \mid a_n < b_n\}$, $B := \{n \in \mathbb{N} \mid a_n = b_n\}$, and $C = \{n \in \mathbb{N} | \ a_n > b_n\}$. Then $\mathbb{N} = A \cup B \cup C$ and A, B, C are pairwise disjoint.

Thus, by Exercise 2.14, exactly one of A, B, C is in $\mathcal{U}$, corresponding to the situations $[(a_n)] < [(b_n)]$, $[(a_n)] = [(b_n)]$, and $[(a_n)] > [(b_n)]$. (This is the first time that we have used the *ultra-* assumption.)

Set $\mathbf{0} := [(0, 0, 0, \ldots)]$ and $\mathbf{1} := [(1, 1, 1, \ldots)]$.

Theorem 2.16. $(\mathbb{R}^{\mathcal{U}}, +, \cdot, \mathbf{0}, \mathbf{1}, <)$ *is an ordered field.*

Proof. Checking the ordered field axioms is mainly routine. Let us only verify the existence of multiplicative inverses. Suppose that $[(a_n)] \neq \mathbf{0}$, that is, that $\{n \in \mathbb{N} \mid a_n \neq 0\} \in \mathcal{U}$. Define a new sequence (b_n) by $b_n = a_n$ if $a_n \neq 0$ and $b_n = 1$ if $a_n = 0$. (The choice of 1 is irrelevant here; all that matters is we defined $b_n \neq 0$.) Observe that $[(a_n)] = [(b_n)]$ and $b_n \neq 0$ for all n. Now $[(a_n)] \cdot [(b_n^{-1})] = [(b_n)] \cdot [(b_n^{-1})] = \mathbf{1}$, whence $[(a_n)]$ has a multiplicative inverse. □

We claim that $\mathbb{R}^{\mathcal{U}}$ (for suitable $\mathcal{U}$) will serve as a nonstandard extension. First, in order to verify postulate (NS1), we need $\mathbb{R}$ to be an ordered subfield of $\mathbb{R}^{\mathcal{U}}$. This can't literally be true as $\mathbb{R}$ is not a sub*set* of $\mathbb{R}^{\mathcal{U}}$. Instead, we are going to define an ordered field embedding $d : \mathbb{R} \to \mathbb{R}^{\mathcal{U}}$ (and then, following ordinary mathematical practice, we pretend that $\mathbb{R}$ "is" $d(\mathbb{R}) \subseteq \mathbb{R}^{\mathcal{U}}$, that is, we identify $\mathbb{R}$ with its image $d(\mathbb{R})$). The embedding d is given by $d(r) = [(r, r, r, \ldots)]$. For example, $d(0) = \mathbf{0}$ and $d(1) = \mathbf{1}$. It is straightforward to verify that d is an ordered field embedding, called the *diagonal embedding*. In this sense, (NS1) is satisfied.

What about (NS2)? Here, we need to specify an extra condition on our ultrafilter $\mathcal{U}$. To see this, we first explain why principal ultrafilters are boring:

Exercise 2.17. Prove that the diagonal embedding $d : \mathbb{R} \to \mathbb{R}^{\mathcal{U}}$ is surjective (that is, an isomorphism) if and only if $\mathcal{U}$ is principal.

The whole point of nonstandard extensions is to get new elements, so principal ultrapowers (that is $\mathbb{R}^{\mathcal{U}}$ for $\mathcal{U}$ principal) do not help us achieve the goals of nonstandard analysis. So, we should use nonprincipal ultrafilters. But do they exist?

Theorem 2.18. *There exists a nonprincipal ultrafilter on* $\mathbb{N}$.

Proof. Let $\mathcal{F}$ be the Frechet filter on $\mathbb{N}$. By Zorn's Lemma, there is a maximal filter $\mathcal{U}$ on $\mathbb{N}$ containing $\mathcal{F}$. We leave it as an exercise to the reader to show that $\mathcal{U}$ is a nonprincipal ultrafilter. (Hint: Use Exercise 2.9 to show that $\mathcal{U}$ is an ultrafilter.) □

For the rest of this subsection, fix a nonprincipal ultrafilter $\mathcal{U}$ on $\mathbb{N}$. We can now verify (NS2). For notational reasons, it will be simpler to construct a positive infinite element $\alpha \in \mathbb{R}^{\mathcal{U}}$; then $\epsilon := \frac{1}{\alpha}$ will be the desired positive infinitesimal. We claim that $\alpha := [(1, 2, 3, \ldots)]$ is a positive infinite element. To see this, fix $r \in \mathbb{R}^{>0}$; we need to verify that $d(r) < \alpha$, that is, we need to verify $X := \{n \in \mathbb{N} \mid r < n\} \in \mathcal{U}$. Well, if $X \notin \mathcal{U}$, then $\mathbb{N} \setminus X \in \mathcal{U}$; but

$\mathbb{N} \setminus X$ is finite, whence $\mathcal{U}$ is principal by Exercise 2.13. Consequently $X \in \mathcal{U}$ and α is infinite.

For axiom (NS3), we define nonstandard extensions of sets and functions as follows:

- For $f : \mathbb{R}^n \to \mathbb{R}$, we define $f : (\mathbb{R}^{\mathcal{U}})^n \to \mathbb{R}^{\mathcal{U}}$ by
$$f([(a_m^1)], \ldots, [(a_m^n)]) := [(f(a_m^1, \ldots, a_m^n))].$$
- For $A \subseteq \mathbb{R}^n$, we define $A^* \subseteq (\mathbb{R}^{\mathcal{U}})^n$ by
$$([(a_m^1)], \ldots, [(a_m^n)]) \in A^* \text{ if and only if } \{m \in \mathbb{N} \mid (a_m^1, \ldots, a_m^n) \in A\} \in \mathcal{U}.$$

As before, one must check that these operations are well-defined (we leave this to the reader) and, after identifying $\mathbb{R}$ with $d(\mathbb{R})$, these functions and relations really do "extend" the original functions and relations (again relegated to the lucky reader to verify).

Finally, what about (NS4)? It is here that logic must reenter the picture in some shape or form. (Up until this point, the nonlogician aiming to use nonstandard methods via the ultrapower approach has been content.) Let $\mathcal{L}$ be the first-order language described in the previous subsection. We make $\mathbb{R}^{\mathcal{U}}$ into an $\mathcal{L}$-structure $\mathfrak{R}^{\mathcal{U}}$ by interpreting F_f and P_A as the extensions defined above. Then the precise formulation of (NS4) is the following:

Theorem 2.19 (Łos')**.** *Suppose that $\varphi(v_1, \ldots, v_n)$ is an $\mathcal{L}$-formula and $[(a_m^1)], \ldots, [(a_m^n)] \in \mathbb{R}^{\mathcal{U}}$. Then*
$$\mathfrak{R}^{\mathcal{U}} \models \varphi[\![[(a_m^1)], \ldots, [(a_m^n)]]\!] \text{ if and only if } \{m \in \mathbb{N} \mid \mathfrak{R} \models \varphi[\![a_m^1, \ldots, a_m^n]\!]\} \in \mathcal{U}.$$

Proof. A useful exercise in logic; proceed by induction on the complexity of φ. □

Observe that, as a corollary of Łos' theorem, that $\mathbb{R}^{\mathcal{U}}$ is an ordered field (as these axioms are first-order). The analyst trying to refrain from logic surely avoids Łos' theorem, but in practice, ends up repeatedly verifying its conclusion on a case-by-case basis.

In summary, we obtain:

Theorem 2.20. *$\mathfrak{R}^{\mathcal{U}}$ is a nonstandard universe.*

2.3. **Problems.**

Problem 2.1. Discuss how to make define the nonstandard extension of functions $f : A \to B$ with $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$.

Problem 2.2. Let $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$.

(1) Suppose $f : A \to B$ is 1-1. Show that if $a \in A^* \setminus A$, then $f(a) \in B^* \setminus B$.
(2) Show that A is finite if and only if $A^* = A$.

3. Sequences and series

3.1. **First results about sequences.** OK, so let's start doing some calculus nonstandardly. We start by studying sequences and series. A *sequence* is a function $s : \mathbb{N} \to \mathbb{R}$. We often write $(s_n \mid n \in \mathbb{N})$ or just (s_n) to denote a sequence, where $s_n := s(n)$. By Exercise 2.3, we have a nonstandard extension $s : \mathbb{N}^* \to \mathbb{R}^*$, which we also often denote by $(s_n \mid n \in \mathbb{N}^*)$.

Notation: We write $N > \mathbb{N}$ to indicate $N \in \mathbb{N}^* \setminus \mathbb{N}$.

Definition 3.1. (s_n) *converges to* L, written $(s_n) \to L$ or $\lim_{n\to\infty} s_n = L$, if: for all $\epsilon \in \mathbb{R}^{>0}$, there is $m \in \mathbb{N}$ such that, for all $n \in \mathbb{N}$, if $n \geq m$, then $|s_n - L| < \epsilon$.

We now give our first nonstandard characterization of a standard concept.

Theorem 3.2. *$s_n \to L$ if and only if $s_N \approx L$ for all $N > \mathbb{N}$.*

This theorem lends solid ground to the heuristic expression: $s_n \to L$ if and only if, for *really large* N, s_N is *really close to* L.

Proof. ($\Rightarrow$) Suppose $s_n \to L$. Fix $N > \mathbb{N}$. We want $s_N \approx L$. Fix $\epsilon \in \mathbb{R}^{>0}$. We want $|s_N - L| < \epsilon$. By assumption, there is $m \in \mathbb{N}$ such that

$$\mathfrak{R} \models \forall n \in \mathbb{N}(n \geq m \to |s_n - L| < \epsilon).$$

(Here, we are mixing formal logic with informal notation. If we were being polite, we would write

$$\mathfrak{R} \models \forall n((P_{\mathbb{N}} n \wedge P_{\geq} nm) \to P_{<} F_{|*-*|} F_s n c_L c_\epsilon).$$

After seeing the formal version, hopefully you will forgive our rudeness and allow us to write in hybrid statements as above!) Thus, by the Transfer Principle, $\mathfrak{R}^* \models \forall n \in \mathbb{N}^*(n \geq m \to |s_n - L| < \epsilon)$. Since $m \in \mathbb{N}$ and $N > \mathbb{N}$, we have $N \geq m$. Thus, $|s_N - L| < \epsilon$, as desired.

($\Leftarrow$) We now suppose $s_N \approx L$ for $N > \mathbb{N}$. Fix $\epsilon \in \mathbb{R}^{>0}$. We need $m \in \mathbb{N}$ such that, $n \in \mathbb{N}$ and $n \geq m$ implies $|s_n - L| < \epsilon$. But how are we to find such m? Well, $\mathfrak{R}^*$ knows of such an m (satisfying the $*$-version of the desired condition). Indeed, if $m > \mathbb{N}$, then $n \in \mathbb{N}^*$ and $n \geq m$ implies $n > \mathbb{N}$, whence $s_n \approx L$ and, in particular, $|s_n - L| < \epsilon$. So, $\mathfrak{R}^* \models \exists m \in \mathbb{N}^* \forall n \in \mathbb{N}^*(n \geq m \to |s_n - L| < \epsilon)$. Thus, by the Transfer Principle, $\mathfrak{R} \models \exists m \in \mathbb{N} \forall n \in \mathbb{N}(n \geq m \to |s_n - L| < \epsilon)$, as desired. □

We used transfer in each direction of the previous proof. The first application is often called "Upward Transfer" as a fact from below (in the "real world") was transferred up to the nonstandard world. Similarly, the second application is often called "Downward Transfer" for a similar reason.

Theorem 3.3 (Monotone Convergence)**.** *Let (s_n) be a sequence.*

(1) *Suppose (s_n) is bounded above and nondecreasing. Then (s_n) converges to* $\sup\{s_n \mid n \in \mathbb{N}\}$.

(2) *Suppose* (s_n) *is bounded below and nonincreasing. Then* (s_n) *converges to* $\inf\{s_n \mid n \in \mathbb{N}\}$*..*

Proof. We only prove (1); the proof of (2) is similar. By the previous theorem, it suffices to prove the following

Claim: For any $N > \mathbb{N}$, $s_N \in \mathbb{R}_{\text{fin}}$ and $\text{st}(s_N) = \sup\{s_n \mid n \in \mathbb{N}\}$.
Proof of Claim: Let b be an upper bound for (s_n), so

$$\mathfrak{R} \models \forall n \in \mathbb{N}(s_1 \leq s_n \leq b).$$

Thus, by transfer, $\mathfrak{R}^* \models \forall n \in \mathbb{N}^*(s_1 \leq s_n \leq b)$. Thus, $s_N \in \mathbb{R}_{\text{fin}}$ for all $N \in \mathbb{N}^*$. Fix $N > \mathbb{N}$ and let $L := \text{st}(s_N)$. We need to show $L = \sup\{s_n \mid n \in \mathbb{N}\}$. By the transfer principle applied to the nondecreasing assumption, $s_n \leq s_N$ for all $n \in \mathbb{N}$, whence $s_n \leq \text{st}(s_N) = L$ for all $n \in \mathbb{N}$. Why is L the least upper bound of $\{s_n \mid n \in \mathbb{N}\}$? Let r be an upper bound for $\{s_n \mid n \in \mathbb{N}\}$. Then $\mathfrak{R} \models \forall n \in \mathbb{N}(s_n \leq r)$. By the transfer principle, $\mathfrak{R}^* \models \forall n \in \mathbb{N}^*(s_n \leq r)$; in particular, $s_N \leq r$, whence $L = \text{st}(s_N) \leq r$. □

Corollary 3.4. *Suppose* $c \in (0, 1)$*. Then* $c^n \to 0$*.*

Proof. $(c^n \mid n \in \mathbb{N})$ is bounded below (by 0) and nonincreasing since $c \in (0, 1)$. Thus, $c^n \to L$ for some L. Fix $N > \mathbb{N}$; by Theorem 3.2, $c^N \approx L$ and $c^{N+1} \approx L$. So

$$L \approx c^{N+1} = c \cdot c^n \approx c \cdot L.$$

Since L, cL are both standard numbers, we have $cL = L$. Since $c \neq 1$, we have $L = 0$. □

Definition 3.5. (s_n) is *bounded* if there is $b \in \mathbb{R}$ such that $|s_n| < b$ for all $n \in \mathbb{N}$.

Proposition 3.6. (s_n) *is bounded if and only if* $s_N \in \mathbb{R}_{\text{fin}}$ *for all* $N \in \mathbb{N}^*$*.*

Proof. The $(\Rightarrow)$ direction follows immediately from transfer. For the converse, suppose that $s_N \in \mathbb{R}_{\text{fin}}$ for all $N \in \mathbb{N}^*$. Fix $M > \mathbb{N}$. Then $|s_N| \leq M$ for all $N \in \mathbb{N}^*$, so $\mathfrak{R}^* \models (\exists M \in \mathbb{N}^*)(\forall n \in \mathbb{N}^*)(|s_N| \leq M)$. Applying the transfer principle to this statement yields a standard bound for (s_n). □

Definition 3.7. (s_n) is a *Cauchy sequence* if for all $\epsilon \in \mathbb{R}^{>0}$, there is $k \in \mathbb{N}$ such that, for all $m, n \in \mathbb{N}$, if $m, n \geq k$, then $|s_m - s_n| < \epsilon$.

Lemma 3.8. *If* (s_n) *is Cauchy, then* (s_n) *is bounded.*

Proof. We use the nonstandard criteria for boundedness. Fix $N > \mathbb{N}$; we must show $s_N \in \mathbb{R}_{\text{fin}}$. Set $A := \{n \in \mathbb{N}^* \mid |s_n - s_N| < 1\}$, so A is a definable set in $\mathfrak{R}^*$. Since (s_n) is Cauchy, there is $k \in \mathbb{N}$ such that, for $m, p \in \mathbb{N}$, if $m, p \geq k$, then $|s_m - s_p| < 1$. Applying the transfer principle to this last statement, if $n > \mathbb{N}$, then $|s_n - s_N| < 1$. Thus, $\mathbb{N}^* \setminus \mathbb{N} \subseteq A$. Since $\mathbb{N}^* \setminus \mathbb{N}$ is not definable, we have $\mathbb{N}^* \setminus \mathbb{N} \subsetneq A$. In other words, there is $n \in A \cap \mathbb{N}$. Since $s_n \in \mathbb{R}$ for $n \in \mathbb{N}$, we see that $s_N \in \mathbb{R}_{\text{fin}}$. □

Exercise 3.9. (s_n) is Cauchy if and only if, for all $M, N > \mathbb{N}$, $s_M \approx s_N$.

Proposition 3.10. *(s_n) converges if and only if (s_n) is Cauchy.*

Proof. The $(\Rightarrow)$ direction is an easy exercise, so we only prove the $(\Leftarrow)$ direction. Suppose (s_n) is Cauchy; then (s_n) is bounded. Fix $M > \mathbb{N}$; then $s_M \in \mathbb{R}_{\text{fin}}$. Set $L := \text{st}(s_M)$. If $N > \mathbb{N}$ is another infinite natural number, then by the previous exercise, $s_N \approx s_M$, so $s_N \approx L$. Thus, (s_n) converges to L. □

3.2. **Cluster points.** If (s_n) is a sequence and $L \in \mathbb{R}$, then we say that L is a *cluster point* of (s_n) if, for each $\epsilon \in \mathbb{R}^{>0}$, the interval $(L - \epsilon, L + \epsilon)$ contains infinitely many s_n's. It will be useful for us to write this in another way: L is a cluster point of (s_n) if and only if, for every $\epsilon \in \mathbb{R}^{>0}$, for every $m \in \mathbb{N}$, there is $n \in \mathbb{N}$ such that $n \geq m$ and $|s_n - L| < \epsilon$.

We can also recast this notion in terms of subsequences. A *subsequence* of (s_n) is a sequence (t_k) such that there is an increasing function $\alpha : \mathbb{N} \to \mathbb{N}$ satisfying $t_k = s_{\alpha(k)}$. We often write (s_{n_k}) for a subsequence of (s_n), where $n_k := \alpha(k)$.

Exercise 3.11. L is a cluster point of (s_n) if and only if there is a subsequence (s_{n_k}) of (s_n) that converges to L.

Recall that (s_n) converges to L if $s_N \approx L$ *for all* $N > \mathbb{N}$. Changing the quantifier "for all" to "there exists" gives us the notion of cluster point:

Proposition 3.12. *L is a cluster point of (s_n) if and only if there is $N > \mathbb{N}$ such that $s_N \approx L$.*

Proof. $(\Rightarrow)$: Apply the transfer principle to the definition of cluster point. Fix $\epsilon \in \mu^{>0}$ and $M > \mathbb{N}$. Then there is $N \in \mathbb{N}^*$, $N \geq M$, such that $|s_N - L| < \epsilon$. This is the desired N since $N > \mathbb{N}$ and ϵ is infinitesimal.
$(\Leftarrow)$: Fix $N > \mathbb{N}$ such that $s_N \approx L$. Fix $\epsilon \in \mathbb{R}^{>0}$, $m \in \mathbb{N}$. Then

$$\mathfrak{R}^* \models (\exists n \in \mathbb{N}^*)(n \geq m \wedge |s_n - L| < \epsilon);$$

indeed, N witnesses the truth of this quantifier. It remains to apply transfer to this statement. □

We immediately get the famous:

Corollary 3.13 (Bolzano-Weierstraß)**.** *Suppose that (s_n) is bounded. Then (s_n) has a cluster point.*

Proof. Fix $N > \mathbb{N}$. Then since (s_n) is bounded, $s_N \in \mathbb{R}_{\text{fin}}$. Let $L = \text{st}(s_N)$; then L is a cluster point of (s_n) by the last proposition. □

Suppose $s = (s_n)$ is a bounded sequence. Let $C(s)$ denote the set of cluster points of s. Then, by the previous proposition, we have

$$C(s) = \{L \in \mathbb{R} \mid s_N \approx L \text{ for some } N > \mathbb{N}\}.$$

Exercise 3.14. $C(s)$ is a bounded set.

We may thus make the following definitions:

Definition 3.15.

(1) $\limsup s_n := \sup C(s)$;
(2) $\liminf s_n := \inf C(s)$.

It turns out that $\limsup s_n$ and $\liminf s_n$ are themselves cluster points of s, that is, these suprema and infima are actually max and min:

Proposition 3.16. *$\limsup s_n, \liminf s_n \in C(s)$.*

Proof. We only prove this for $\limsup s_n$; the proof for $\liminf s_n$ is similar. For simplicity, set $r := \limsup s_n$. Fix $\epsilon > 0$ and $m \in \mathbb{N}$. We need to find $n \in \mathbb{N}$ such that $|r - s_n| < \epsilon$. Since $r := \sup C(s)$, there is $L \in C(s)$ such that $r - \epsilon < L \leq r$. Take $N > \mathbb{N}$ such that $L = \text{st}(s_N)$. Then $r - \epsilon < s_N < r + \epsilon$. Now apply transfer. □

Theorem 3.17. *$(s_n) \to L$ if and only if $\limsup s_n = \liminf s_n = L$.*

Proof. $(s_n) \to L$ if and only if $\text{st}(s_N) = L$ for all $N > \mathbb{N}$ if and only if $C(s) = \{L\}$ if and only if $\limsup s_n = \liminf s_n = L$. □

There is an alternate description of $\limsup s_n$ and $\liminf s_n$ in terms of tail sets which we now explain. For $n \in \mathbb{N}$, set $T_n := \{s_n, s_{n+1}, s_{n+2}, \ldots\}$, the n^{th} *tailset of* s. Set $S_n := \sup T_n$. Since $T_{n+1} \subseteq T_n$, we have $S_{n+1} \leq S_n$, whence (S_n) is a nonincreasing sequence. Since (s_n) is bounded, so is (S_n). Thus, by the Monotone Convergence Theorem, (S_n) converges to $\inf S_n$.

Theorem 3.18. *$\limsup s_n = \lim S_n = \inf S_n$.*

Before we can prove this result, we need a preliminary result:

Lemma 3.19. *Let $\epsilon \in \mathbb{R}^{>0}$. Then there is $m \in \mathbb{N}$ such that, for all $n \in \mathbb{N}$, if $n \geq m$, then $s_n < \limsup s_n + \epsilon$.*

Proof. Fix $N > \mathbb{N}$. Then $\text{st}(s_N) \leq \limsup s_n$, so $s_N < \limsup s_n + \epsilon$. Thus, any $M > \mathbb{N}$ witness that

$$\mathfrak{R}^* \models (\exists M \in \mathbb{N}^*)(\forall N \in \mathbb{N}^*)(N \geq M \Rightarrow s_N < \limsup s_n + \epsilon).$$

Now apply the transfer principle. □

Proof. (of Theorem 3.18) Fix $m \in \mathbb{N}$. Then by definition, if $n \in \mathbb{N}$ and $n \geq m$, then $s_n \leq S_m$. By transfer, this holds for all $n \in \mathbb{N}^*$ with $n \geq m$. Take $N > \mathbb{N}$ such that $\limsup s_n = \text{st}(s_N)$. Then $s_N \leq S_m$, so $\limsup s_n = \text{st}(s_N) \leq S_m$. Since $m \in \mathbb{N}$ is arbitrary, we see that $\limsup s_n \leq \inf S_n$. Now suppose, towards a contradiction, that $\limsup s_n < \inf S_n$. Choose $\epsilon \in \mathbb{R}^{>0}$ such that $\limsup s_n + \epsilon < \inf S_n$. By the previous lemma, there is $m \in \mathbb{N}$ such that, for all $n \geq m$, $s_n < \limsup s_n + \epsilon$. Thus, $S_m \leq \limsup s_n + \epsilon < \inf S_n$, a contradiction. □

Exercise 3.20. State and prove an analog of Theorem 3.18 for $\liminf s_n$.

3.3. **Series.** Recall that an infinite series $\sum_{k=0}^{\infty} a_k$ is *convergent* if the sequence of *partial sums* (s_n) converges, where $s_n := \sum_{k=0}^{n} a_k$. In this case, we write $\sum_{k=0}^{\infty} a_k = L$ if (s_n) converges to L. It is useful to observe that $\sum_{k=m}^{n} a_k = s_n - s_{m-1}$.

Some notation: If $N > \mathbb{N}$, then we have an element s_N of the extended sequence. We use this to define $\sum_{k=0}^{N} a_k$, that is, we set $\sum_{k=0}^{N} a_k := s_N$. Similarly, if $M > \mathbb{N}$ is also infinite and $M \leq N$, we set $\sum_{k=M}^{N} a_k := s_N - s_{M-1}$.

Proposition 3.21.

(1) *$\sum_{k=0}^{\infty} a_k = L$ if and only if $\sum_{k=0}^{N} a_k \approx L$ for all $N > \mathbb{N}$.*
(2) *(Cauchy Criteria) $\sum_{k=0}^{\infty} a_k$ converges if and only if $\sum_{k=M}^{N} a_k \approx 0$ for all $M, N > \mathbb{N}$ with $M \leq N$.*

Proof. (1) follows immediately from the nonstandard characterization of convergent sequences, while (2) follows immediately from the fact that a sequence converges if and only if it is Cauchy together with the nonstandard characterization of Cauchy sequence given in Exercise 3.9. □

Corollary 3.22. *If $\sum a_k$ converges, then $a_k \to 0$ as $k \to \infty$.*

Proof. Let $N > \mathbb{N}$. Then $a_N = s_N - s_{N-1} \approx 0$ by part (2) of the previous proposition. □

The usual warning is relevant here: the converse of the above corollary is not true. Indeed, $\sum 1/k$ diverges even though $1/k \to 0$ as $k \to \infty$.

Definition 3.23. $\sum a_k$ *converges absolutely* if $\sum |a_k|$ converges.

Exercise 3.24. If $\sum a_k$ converges absolutely, then it converges.

Theorem 3.25 (Ratio Test)**.** *Suppose $a_k \neq 0$ for all k.*

(1) *If $\limsup \frac{|a_{k+1}|}{|a_k|} < 1$, then $\sum a_k$ converges absolutely.*
(2) *If $\liminf \frac{|a_{k+1}|}{|a_k|} > 1$, then $\sum a_k$ diverges.*

Proof. (1) Choose $N > \mathbb{N}$ such that $\limsup \frac{|a_{k+1}|}{|a_k|} = \operatorname{st}(\frac{|a_{N+1}|}{|a_N|})$. Fix $c \in \mathbb{R}$ satisfying $\limsup \frac{|a_{k+1}|}{|a_k|} < c < 1$. If $M > \mathbb{N}$, then $\operatorname{st}(\frac{|a_{M+1}|}{|a_M|}) \leq \limsup \frac{|a_{k+1}|}{|a_k|} < c$, whence $|\frac{a_{M+1}}{a_M}| < c$ and hence $|a_{M+1}| < c|a_M|$. By transfer, there is $m \in \mathbb{N}$ such that, for all $k \geq m$, $|a_{k+1}| < c|a_k|$. By induction, $|a_{k+m}| < c^k |a_m|$ for all $k \in \mathbb{N}$, whence $\sum_{k+m}^{k+n} |a_j| \leq (\sum_m^n c^i)|a_m|$ for all $m, n \in \mathbb{N}$ with $m \leq n$. Fix $M, N > \mathbb{N}$ with $M \leq N$. We will show that $\sum_M^N |a_j| \approx 0$. Indeed, $\sum_M^N |a_j| \leq (\sum_{M-m}^{N-m} c^j)|a_m|$. Use the fact that the geometric series $\sum c^i$ converges and $M - m, N - m > \mathbb{N}$ to conclude that $\sum_{M-m}^{N-m} c^j \approx 0$.

(2) Choose $N > \mathbb{N}$ such that $\liminf \frac{|a_{k+1}|}{|a_k|} = \operatorname{st}(\frac{|a_{N+1}|}{|a_N|})$. For any $M > \mathbb{N}$, we have $\operatorname{st}(\frac{|a_{M+1}|}{|a_M|}) \geq \operatorname{st}(\frac{|a_{N+1}|}{|a_N|}) > 1$, whence $|a_{M+1}| > |a_M|$. Thus, $\mathfrak{R}^* \models$

$(\exists N \in \mathbb{N}^*)(\forall M \in \mathbb{N}^*)(M \geq N \rightarrow |a_{M+1}| \geq |a_M|)$. Apply transfer to this statement, yielding $n \in \mathbb{N}$ such that $|a_{m+1}| \geq |a_m|$ for all $m \geq n$. By induction, this yields $|a_k| > |a_n|$ for all $k > n$. Thus, by transfer again, if $K > \mathbb{N}$, then $|a_K| > |a_n| > 0$, so $a_K \not\approx 0$, whence $a_k \not\to 0$ and thus $\sum a_k$ diverges. □

3.4. **Problems.**

Problem 3.1. Suppose that $s_n \to L$ and $t_n \to M$. Show that:

(1) $(s_n \pm t_n) \to L \pm M$;
(2) $(cs_n) \to cL$ for any $c \in \mathbb{R}$;
(3) $(s_n t_n) \to LM$;
(4) $(\frac{s_n}{t_n}) \to \frac{L}{M}$ if $M \neq 0$;
(5) If $s_n \leq t_n$ for all $n \in \mathbb{N}$, then $L \leq M$.

Problem 3.2.

(1) Let (s_n) be a sequence. Show that (s_n) is bounded above if and only if s_N is *not* positive infinite for all $N > \mathbb{N}$. Likewise, show that (s_n) is bounded below if and only if s_N is *not* negative infinite for all $N > \mathbb{N}$.
(2) Say that (s_n) *converges to positive infinity* if for every $r \in \mathbb{R}$, there is $n \in \mathbb{N}$ such that, for all $m \in \mathbb{N}$, if $m \geq n$, then $s_m > r$. Show that (s_n) converges to positive infinity if and only if for every $N > \mathbb{N}$, s_N is positive infinite. Define what it means for (s_n) to *converge to negative infinity* and give an analogous nonstandard description.

Problem 3.3. Suppose that $\sum_0^\infty a_i = L$ and $\sum_1^\infty b_i = M$. Show that:

(1) $\sum_0^\infty (a_i \pm b_i) = L \pm M$;
(2) $\sum_0^\infty (ca_i) = cL$ for any $c \in \mathbb{R}$.

Problem 3.4. Let $c \in \mathbb{R}$. Recall the identity

$$1 + c + c^2 + \cdots + c^n = \frac{1 - c^{n+1}}{1 - c}.$$

Use this identity to show that $\sum_0^\infty c^i$ converges if $|c| < 1$. Conclude that $\sum_m^n c^i$ is infinitesimal if $|c| < 1$ and $m, n > \mathbb{N}$ are such that $m \leq n$.

Problem 3.5. Suppose that $(r_n), (s_n), (t_n)$ are three sequences such that $r_n \leq s_n \leq t_n$ for all $n \in \mathbb{N}$. Further suppose that $L \in \mathbb{R}$ is such that $r_n \to L$ and $t_n \to L$. Show that $s_n \to L$. (This is often referred to as the *Squeeze Theorem.*)

Problem 3.6. Suppose that s, t are bounded sequences. Show that

$$\limsup(s + t) \leq \limsup s + \limsup t.$$

Problem 3.7. Suppose that $a_i \geq 0$ for all $i \in \mathbb{N}$. Prove that the following are equivalent:

(1) $\sum_0^\infty a_i$ converges;

(2) $\sum_0^N a_i \in \mathbb{R}_{\text{fin}}$ for *all* $N > \mathbb{N}$;
(3) $\sum_0^N a_i \in \mathbb{R}_{\text{fin}}$ for *some* $N > \mathbb{N}$.

(Hint: Use the Monotone Convergence Theorem.)

Problem 3.8. Let (s_n) be a sequence. Show that (s_n) is Cauchy if and only if for every $M, N > \mathbb{N}$, we have $s_M \approx s_N$.

Problem 3.9 (Advanced)**.**

(1) Let $\sum_0^\infty a_i$ and $\sum_0^\infty b_i$ be two series, where $a_n, b_n \geq 0$ for all $i \in \mathbb{N}$. Suppose that $a_n \leq b_n$ for all $n > \mathbb{N}$. Further suppose that $\sum_0^\infty b_i$ converges. Show that $\sum_0^\infty a_i$ converges.
(2) Show that the condition "$a_n \leq b_n$ for all $n > \mathbb{N}$" is equivalent to the condition "there exists $k \in \mathbb{N}$ such that $a_n \leq b_n$ for all $n \in \mathbb{N}$ with $n \geq k$," i.e. (b_n) *eventually dominates* (a_n).

The result established in this problem is usually called the *Comparison Test*.

4. Continuity

4.1. **First results about continuity.** Let $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$ a function, and $c \in A$. We say that f is *continuous at* c if: for all $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $x \in \mathbb{R}$, if $x \in A$ and $|x-c| < \delta$, then $|f(x)-f(x)| < \epsilon$. We say that f is *continuous* if f is continuous at c for all $c \in A$. Here is the nonstandard characterization of continuity:

Theorem 4.1. *Suppose $f : A \to \mathbb{R}$ and $c \in A$. The following are equivalent:*

(1) *f is continuous at c;*
(2) *if $x \in A^*$ and $x \approx c$, then $f(x) \approx f(c)$;*
(3) *there is $\delta \in \mu^{>0}$ such that, for all $x \in A^*$, if $|x - c| < \delta$, then $f(x) \approx f(c)$.*

The equivalence between (1) and (2) backs up our usual heuristic that f is continuous at c if, for all $x \in A$ *really close to* c, we have $f(x)$ is *really close to* $f(c)$.

Proof. (1) $\Rightarrow$ (2): Suppose that $x \approx c$; we want $f(x) \approx f(c)$. Fix $\epsilon \in \mathbb{R}^{>0}$; we want $|f(x) - f(c)| < \epsilon$. By (1), there is $\delta \in \mathbb{R}^{>0}$ such that

$$\mathfrak{R} \models \forall x((x \in A \wedge |x - c| < \delta) \to |f(x) - f(c)| < \epsilon).$$

Applying transfer to this statement and realizing that $x \approx c$ implies $|x-c| < \delta$ yields that $|f(x) - f(c)| < \epsilon$.

(2) $\Rightarrow$ (3) follows by taking $\delta \in \mu^{>0}$ arbitrary.

(3) $\Rightarrow$ (1): Fix δ as in (3). Fix $\epsilon \in \mathbb{R}^{>0}$. Since $x \in A^*$ and $|x - c| < \delta$ implies $f(x) \approx f(c)$, and, in particular, $|f(x) - f(c)| < \epsilon$, we have

$$\mathfrak{R}^* \models (\exists \delta \in \mathbb{R}^{>0})(\forall x \in A)(|x - c| < \delta \to |f(x) - f(c)| < \epsilon).$$

Apply transfer. $\square$

Example 4.2. Since $|\sin x| \leq |x|$ for x small, we see that, by transfer, $\sin \epsilon \in \mu$ for $\epsilon \in \mu$. (In other words, sin is continuous at 0.) A similar argument, shows that cos is continuous at 0, that is, $\cos \epsilon \approx 1$ for $\epsilon \in \mu$. Using this and the transfer of the usual trigonometric identities, we can prove that sin is continuous: if $c \in \mathbb{R}$ and $x \approx c$, write $x = c + \epsilon$ for $\epsilon \in \mu$. Then

$$\sin x = \sin(c + \epsilon) = \sin c \cos \epsilon + \cos c \sin \epsilon \approx \sin c \cdot 1 + \cos c \cdot 0 = \sin c.$$

Example 4.3. Let

$$f(x) = \begin{cases} \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then f is *not* continuous at 0. Indeeed, let $N > \mathbb{N}$ be odd and set $x := \frac{2}{N\pi} \approx 0$. Then $f(x) = \sin \frac{N\pi}{2} = 1 \not\approx f(0)$.

However, the function g defined by

$$g(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is continuous at 0. Indeed, suppose $x \approx 0$, $x \neq 0$. Then since $|\sin(1/x)| \leq 1$, $|g(x)| = |x||\sin(1/x)| \approx 0$.

How about the usual connection between continuity and limits? We say that $\lim_{x \to c} f(x) = L$ if: for all $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $x \in A$, if $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$.

Exercise 4.4. Prove that $\lim_{x \to c} f(x) = L$ if and only if, for all $x \in A^*$, if $x \approx c$ but $x \neq c$, then $f(x) \approx L$.

Corollary 4.5. *f is continuous at c if and only if $\lim_{x \to c} f(x) = f(c)$.*

Proposition 4.6. *Suppose that f is continuous at c and g is continuous at $f(c)$. Then $g \circ f$ is continuous at c.*

Proof. If $x \approx c$, then $f(x) \approx f(c)$, whence $g(f(x)) \approx g(f(c))$. $\square$

The following result is fundamental:

Theorem 4.7 (Intermediate Value Theorem). *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous. Let d be a value strictly in between $f(a)$ and $f(b)$. Then there is $c \in (a, b)$ such that $f(c) = d$.*

The nonstandard proof of the Intermediate Value Theorem will be our first example of using so-called "hyperfinite methods;" in this case, we will be using hyperfinite partitions. The idea is to partition the interval $[a, b]$ into subintervals of width $\frac{1}{N}$ for $N > \mathbb{N}$. Then, logically, this partition will behave as if it were finite; in particular, we will be able to detect "the last time" some particular phenomenon happens. However, since we are taking infinitesimal steps, the change in function value at this turning point will be infinitesimal. Here are the precise details:

Proof. Without loss of generality, $f(a) < f(b)$, so $f(a) < d < f(b)$. Define a sequence (s_n) as follows: for $n > 0$, let $\{p_0, \ldots, p_n\}$ denote the partition of $[a, b]$ into n equal pieces of width $\frac{b-a}{n}$, so $p_0 = a$ and $p_n = b$. Since $f(p_0) < d$, we are entitled to define the number $s_n := \max\{p_k \mid f(p_k) < d\}$, so p_k is the "last time" that $f(p_k) < d$. Observe that $s_n < b$.

We now fix $N > \mathbb{N}$ and claim that $c := \mathrm{st}(s_N) \in [a, b]$ is as desired, namely, that $f(c) = d$. (Note that $s_N \in [a, b]$, whence $\mathrm{st}(s_N)$ is defined.) Indeed, by transfer, $s_N < b$, whence $s_N + \frac{b-a}{N} \leq b$. Again, by transfer, $f(s_N) < d \leq f(s_N + \frac{b-a}{N})$. However, $s_N + \frac{b-a}{N} \approx s_N \approx c$. Since f is continuous at c, we have

$$f(c) \approx f(s_N) < d \leq f(s_N + \frac{b-a}{N}) \approx f(c),$$

whence $f(c) \approx d$. Since $f(c), d \in \mathbb{R}$, we get that $f(c) = d$. $\square$

The next fundamental result is also proven using hyperfinite partitions.

Theorem 4.8 (Extreme Value Theorem)**.** *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous. Then there are $c, d \in [a, b]$ such that $f(c) \leq f(x) \leq f(d)$ for all $x \in [a, b]$. (In other words, f achieves it maximum and minimum.)*

Proof. We will only prove the existence of the maximum of f. Once again, let $\{p_0, p_1, \ldots, p_n\}$ denote the partition of $[a, b]$ into n equal pieces. This time, we define s_n to be some p_k such that $f(p_j) \leq f(p_k)$ for all $j = 0, \ldots, n$. Fix $N > \mathbb{N}$ and set $d := \mathrm{st}(s_N) \in [a, b]$. We claim that f achieves its maximum at d. (Please appreciate the beauty of this claim: We are partitioning $[a, b]$ into hyperfinitely many pieces, looking for where the function achieves its maximum on this hyperfinite set, and claiming that this element is infinitely close to where the function achieves its maximum on $[a, b]$. Magic!)

Fix $x \in [a, b]$. We need $f(x) \leq f(d)$. First, we need to "locate" x in our hyperfinite partition. Since

$$\mathfrak{R} \models (\forall n \in \mathbb{N})(\exists k \in \mathbb{N})(0 \leq k < n \wedge a + \frac{k(b-a)}{n} \leq x \leq a + \frac{(k+1)(b-a)}{n}),$$

by transfer, we can find $k \in \mathbb{N}^*$, $0 \leq k < N$, such that $a + \frac{k(b-a)}{N} \leq x \leq a + \frac{(k+1)(b-a)}{N}$. Since the interval $[a + \frac{k(b-a)}{N}, a + \frac{(k+1)(b-a)}{N}]$ has infinitesimal width $\frac{b-a}{N}$ and f is continuous at x, we see that $f(a + \frac{k(b-a)}{N}) \approx f(x) \approx f(a + \frac{(k+1)(b-a)}{N})$. However, by transfer, $f(a + \frac{k(b-a)}{N}) \leq f(s_N)$. Since $f(s_N) \approx f(d)$ (since f is continuous at d), it follows that $f(x) \leq f(d)$. $\square$

4.2. **Uniform continuity.** Recall that $f : A \to \mathbb{R}$ is continuous if:

$$(\forall x \in A)(\forall \epsilon \in \mathbb{R}^{>0})(\exists \delta \in \mathbb{R}^{>0})(\forall y \in A)(|x - y| < \delta \to |f(x) - f(y)| < \epsilon).$$

The crucial point here is that a particular δ (a priori) depends *both* on the given ϵ *and* the location x. Uniform continuity is what one gets when one demands that the δ depend only on the given ϵ and not on the location

x. In other words, it is what one gets when one slides the first universal quantifier in the above display over to the spot after the $\exists\delta$. More formally:

Definition 4.9. $f : A \to \mathbb{R}$ is *uniformly continuous* if, for all $\epsilon \in \mathbb{R}^{>0}$, there exists $\delta \in \mathbb{R}^{>0}$ such that, for all $x, y \in A$, if $|x-y| < \delta$, then $|f(x)-f(y)| < \epsilon$.

Using purely symbolic language as above, we see that $f : A \to \mathbb{R}$ is uniformly continuous if and only if:

$$(\forall \epsilon \in \mathbb{R}^{>0})(\exists \delta \in \mathbb{R}^{>0})(\forall x, y \in A)(|x - y| < \delta \to |f(x) - f(y)| < \epsilon).$$

While uniform continuity is no more difficult to state than ordinary continuity (as it results from a simple permutation of quantifiers), it is often difficult for students to first digest. For this reason, uniform continuity is perhaps one of the best examples of elucidating a standard concept by nonstandard means. Here is the nonstandard equivalence:

Proposition 4.10. *$f : A \to \mathbb{R}$ is uniformly continuous if and only if, whenever $x, y \in A^*$ are such that $x \approx y$, then $f(x) \approx f(y)$.*

Please make sure that you see how this is different from ordinary continuity. Indeed, $f : A \to \mathbb{R}$ is continuous if and only if, whenever $x, y \in A^*$ are such that $x \approx y$, *and at least one of x and y are standard*, then $f(x) \approx f(y)$. Thus, uniform continuity demands continuity for the extended part of A as well.

Exercise 4.11. Prove Proposition 4.10.

Example 4.12. It is now easy to see why $f : (0, 1] \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ is not uniformly continuous. Indeed, fix $N > \mathbb{N}$. Then $\frac{1}{N}, \frac{1}{N+1} \in (0, 1]^*$, $\frac{1}{N} \approx \frac{1}{N+1}$, but $N = f(\frac{1}{N}) \not\approx f(\frac{1}{N+1}) = N + 1$. Note that this calculation does not contradict the fact that f is continuous. Indeed, since 0 is not in the domain of f, we would never be calculating $f(x)$ for infinitesimal x when verifying continuity.

Recall our above heuristic, namely that $f : A \to \mathbb{R}$ is uniformly continuous when f is continuous on the "extended part" of A. But sometimes A doesn't have an extended part. For example, if $A = [a, b]$, then A^* is such that every element is infinitely close to an element of A. It is for this reason that continuous functions whose domains are closed, bounded intervals are *automatically* uniformly continuous:

Theorem 4.13. *Suppose $f : [a, b] \to \mathbb{R}$ is continuous. Then f is uniformly continuous.*

Proof. Suppose $x, y \in [a, b]^*$, $x \approx y$. Then $c := \mathrm{st}(x) = \mathrm{st}(y) \in [a, b]$. Since f is continuous at c, we have that $f(x) \approx f(c) \approx f(y)$. $\square$

Please compare the proof of the previous theorem with the usual standard proof. In particular, compare the lengths of these proofs!

4.3. **Sequences of functions.** We now would like to consider sequences of functions and different types of convergence of such sequences. For each $n \in \mathbb{N}$, let $f_n : A \to \mathbb{R}$ be a function. Also, let $f : A \to \mathbb{R}$ be another function.

Definition 4.14. We say that f_n *converges to f pointwise* if: for each $x \in A$, the sequence of numbers $(f_n(x))$ converges to the number $f(x)$.

In other words: f_n converges to f pointwise if and only if:

$$(\forall x \in A)(\forall \epsilon \in \mathbb{R}^{>0})(\exists m \in \mathbb{N})(\forall n \in \mathbb{N})(n \geq m \to |f_n(x) - f(x)| < \epsilon).$$

We would like to get a nonstandard characterization of this concept. First, how should we define the extended sequence (f_n)? Clearly, for each $n \in \mathbb{N}$, we get an extended function $f_n : A^* \to \mathbb{R}^*$. But what about infinite N? Here's, the trick: define a function $F : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ by $F(n, x) = f_n(x)$. We thus have a nonstandard extension $F : \mathbb{N}^* \times \mathbb{R}^* \to \mathbb{R}^*$. We now define, for $n \in \mathbb{N}^*$, $f_n : A^* \to \mathbb{R}^*$ by $f_n(x) = F(n, x)$. We should be a bit worried at this point as, for standard $n \in \mathbb{N}$, we have used $f_n : A^* \to \mathbb{R}^*$ to denote two, possibly different functions. Thankfully, by transfer, there is no issue here: we leave it to the reader to check that, for standard $n \in \mathbb{N}$, both uses of $f_n : A^* \to \mathbb{R}^*$ agree. And now we have extended the sequence (f_n) by defining $f_n : A^* \to \mathbb{R}^*$ for $n \in \mathbb{N}^* \setminus \mathbb{N}$.

Exercise 4.15. Prove that f_n converges to f pointwise on A if and only if, for all $x \in A$, for all $N > \mathbb{N}$, $f_N(x) \approx f(x)$.

Once again, we will define the uniform concept by moving the quantifiers and asking the m in the above definition of pointwise convergence to depend only on the ϵ and not on the location x.

Definition 4.16. We say that f_n *converges uniformly to f* if, for all $\epsilon \in \mathbb{R}^{>0}$, there exists $m \in \mathbb{N}$, such that, for all $n \in \mathbb{N}$, if $n \geq m$, then, for all $x \in A$, $|f_n(x) - f(x)| < \epsilon$.

In symbols: f_n converges uniformly to f if and only if:

$$(\forall \epsilon \in \mathbb{R}^{>0})(\exists m \in \mathbb{N})(\forall n \in \mathbb{N})(n \geq m \to (\forall x \in A)(|f_n(x) - f(x)| < \epsilon)).$$

The way to visualize uniform convergence is the following: for a given ϵ, draw the strip around f of radius ϵ; then f_n converges uniformly to f if and only if, eventually, all of the functions f_n live in the strip of radius ϵ around f.

Here's the nonstandard equivalence; notice the similarity in the nonstandard comparison between continuity and uniform continuity.

Proposition 4.17. *f_n converges uniformly to f if and only if, for all $x \in A^*$, for all $N > \mathbb{N}$, $f_N(x) \approx f(x)$.*

Exercise 4.18. Prove Proposition 4.17.

Example 4.19. Consider $f_n : [0,1] \to \mathbb{R}$ given by $f_n(x) = x^n$. Let $f : [0,1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0 & \text{if } x \neq 1 \\ 1 & \text{if } x = 1. \end{cases}$$

Then f_n converges pointwise to f. However, f_n does not converge uniformly to f. Indeed, fix $x \in [0,1]^*$ such that $x < 1$ but $x \approx 1$. We will find $N > \mathbb{N}$ such that $f_N(x) = x^N \not\approx 0 = f(x)$. To do this, set

$$X = \{n \in \mathbb{N}^* \mid |x^n - 1| < \frac{1}{2}\}.$$

Note that X is definable and, by continuity of f_n for $n \in \mathbb{N}$, we have $\mathbb{N} \subseteq X$. Since $\mathbb{N}$ is not definable, there is $N \in X \backslash \mathbb{N}$. For this N, we have $|x^N - 1| < \frac{1}{2}$, whence $x^N \not\approx 0$.

Notice that, in the above example, each f_n was continuous but f was not continuous. This cannot happen if the convergence is uniform:

Theorem 4.20. *Suppose that $f_n : A \to \mathbb{R}$ is continuous for each n and that f_n converges to f uniformly. Then f is continuous.*

Proof. Fix $c \in A$; we want to show that f is continuous at c. Fix $x \in A^*$. By the triangle inequality, we have, for any $n \in \mathbb{N}^*$, that

$$|f(x) - f(c)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)|.$$

We would like to make all three of the quantities in the right hand side of the above inequality infinitesimal. Well, if $n > \mathbb{N}$, then by uniform continuity, we have $f(x) \approx f_n(x)$ and $f(c) \approx f_n(c)$. But we only know that $f_n(x) \approx f_n(c)$ for $n \in \mathbb{N}$. Can we arrange for the middle term to be infinitesimal for infinite n as well? Yes, but we need to be a bit more careful.

Since each f_n is continuous at c, we have that the following is true in $\mathfrak{R}$:

$$(\forall n \in \mathbb{N})(\forall \epsilon \in \mathbb{R}^{>0})(\exists \delta \in \mathbb{R}^{>0})(\forall y \in A)(|y - c| < \delta \to |f_n(y) - f_n(c)| < \epsilon).$$

Apply the transfer principle to this result with $N > \mathbb{N}$ and $\epsilon \in \mu^{>0}$. We then get a $\delta \in (\mathbb{R}^*)^{>0}$ such that, whenever $x \in A^*$ satisfies $|x - c| < \delta$, then $f_N(x) \approx f_N(c)$. Without loss of generality, we may assume that $\delta \in \mu^{>0}$. Thus, for all $x \in A^*$, whenever $|x - c| < \delta$, we have $f(x) \approx f(c)$. By the direction (3) $\Rightarrow$ (1) of Theorem 4.1, we see that f is continuous at c. $\square$

In the above proof, we actually stumbled upon some important concepts.

Definition 4.21. Suppose that $f : A^* \to \mathbb{R}^*$ is a function (not necessarily the nonstandard extension of a standard function). Then f is *internally continuous* or **continuous* if: for all $x \in A^*$, for all $\epsilon \in (\mathbb{R}^*)^{>0}$, there is $\delta \in (\mathbb{R}^*)^{>0}$ such that, for all $y \in A^*$, if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.

The terminology *internally continuous* will make more sense later in these notes when we define internal sets and functions. Notice that being internally continuous is just like ordinary continuity, but with everything decorated by stars. By the transfer principle, if $f : A \to \mathbb{R}$ is continuous, then its nonstandard extension $f : A^* \to \mathbb{R}^*$ is internally continuous. Another example is provided by the above proof: if each element of the sequence f_n is continuous, then $f_n : A^* \to \mathbb{R}^*$ is internally continuous for all $n \in \mathbb{N}^*$.

While we are on the topic, let's define another notion of continuity for nonstandard functions.

Definition 4.22. $f : A^* \to \mathbb{R}^*$ is *S-continuous* if, for all $x, y \in A^*$, if $x \approx y$, then $f(x) \approx f(y)$.

So, for example, if $f : A \to \mathbb{R}$ is a standard function, then f is uniformly continuous if and only if $f : A^* \to \mathbb{R}^*$ is S-continuous.

Here is another nice convergence theorem:

Theorem 4.23 (Dini)**.** *Suppose that $f_n : [a, b] \to \mathbb{R}$ is continuous, $f : [a, b] \to \mathbb{R}$ is continuous and f_n converges pointwise to f. Further suppose that $f_n(x)$ is nonincreasing for each $x \in [a, b]$. Then f_n converges uniformly to f.*

Proof. By subtracting f from all of the functions involved, we may as well assume that f is the zero function. Fix $N > \mathbb{N}$ and $c \in [a, b]^*$. We need to show that $f_N(c) \approx 0$. Fix $n \in \mathbb{N}$. By transfer, $0 \leq f_N(c) \leq f_n(c)$. Set $d := \mathrm{st}(c) \in [a, b]$. Since f_n is continuous, we see that $f_n(c) \approx f_n(d)$. Consequently, $f_N(c) \in \mathbb{R}_{\mathrm{fin}}$. Taking standard parts, we see that

$$0 \leq \mathrm{st}(f_N(c)) \leq \mathrm{st}(f_n(c)) = f_n(d).$$

Since $f_n(d) \to 0$ as $n \to \infty$, we see that $\mathrm{st}(f_N(c)) = 0$. □

In Theorem 4.20, if we assumed that each f_n was uniformly continuous, could we conclude that f was uniformly continuous? Unfortunately, this is not true in general. To explain when this is true, we need to introduce a further notion. First, suppose that each $f_n : A \to \mathbb{R}$ is uniformly continuous. Then, in symbols, this means:

$$(\forall n \in \mathbb{N})(\forall \epsilon \in \mathbb{R}^{>0})(\exists \delta \in \mathbb{R}^{>0})(\forall x, y \in A)(|x-y| < \delta \to |f_n(x)-f_n(y)| < \epsilon).$$

Thus, δ can depend on both the ϵ and the particular function f_n. Here's what happens when we only require the δ to depend on the ϵ:

Definition 4.24. The sequence (f_n) is *equicontinuous* if: for all $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $x, y \in A$, for all $n \in \mathbb{N}$, if $|x - y| < \delta$, then $|f_n(x) - f_n(y)| < \epsilon$.

In other words, (f_n) is equicontinuous if it is uniformly uniformly continuous in the sense that each f_n is uniformly continuous and, for a given ϵ, there is a single δ that witnesses the uniform continuity for each f_n.

Exercise 4.25. Prove that (f_n) is equicontinuous if and only if $f_n : A^* \to \mathbb{R}^*$ is S-continuous for each $n \in \mathbb{N}^*$.

Equicontinuity guarantees that the limit function is uniformly continuous:

Theorem 4.26. *Suppose $f_n : A \to \mathbb{R}$ and $f : A \to \mathbb{R}$ are functions such that f_n converges pointwise to f. If (f_n) is equicontinuous, then f is uniformly continuous.*

Proof. Fix $x, y \in A^*$ with $x \approx y$. We need to prove that $f(x) \approx f(y)$. Fix $\epsilon \in \mu^{>0}$. Choose $n_1, n_2 \in \mathbb{N}^*$ such that, for all $n \in \mathbb{N}^*$:

- if $n \geq n_1$, then $|f_n(x) - f(x)| < \epsilon$;
- if $n \geq n_2$, then $|f_n(y) - f(y)| < \epsilon$.

Such n_1, n_2 exist by the transfer of the pointwise converge assumption. Without loss of generality, we may assume that $n_1, n_2 > \mathbb{N}$. Now fix $N \in \mathbb{N}^*$ with $N > \max(n_1, n_2)$. Then $f_N(x) \approx f(x)$ and $f_N(y) \approx f(y)$. But, by Exercise 4.25, $f_N(x) \approx f_N(y)$. Thus, $f(x) \approx f(y)$, as desired. □

Here's one last equivalence to round out the discussion:

Theorem 4.27. *Suppose each $f_n : [a, b] \to \mathbb{R}$ is continuous and f_n converges pointwise to f. Then f_n converges uniformly to f if and only if (f_n) is equicontinuous.*

Proof. First suppose that f_n converges uniformly to f. Then f is continuous. Since the domain of each f_n and f is a closed, bounded interval, we know that, by Theorem 4.13, that each f_n and f are uniformly continuous. Thus, for $n \in \mathbb{N}$, we have that $f_n : [a, b]^* \to \mathbb{R}^*$ is S-continuous. Now suppose that $N > \mathbb{N}$ and $x, y \in [a, b]^*$ are such that $x \approx y$. By Exercise 4.25, it suffices to prove that $f_N(x) \approx f_N(y)$. However,

$$f_N(x) \approx f(x) \approx f(y) \approx f_N(y),$$

where the first and last $\approx$ are true by uniform convergence, while the middle $\approx$ holds by uniform continuity of f.

Conversely, suppose that (f_n) is equicontinuous; we want to show that f_n converges to f uniformly. To do this, we need to show that, if $N > \mathbb{N}$ and $x \in [a, b]^*$, that $f_N(x) \approx f(x)$. Let $y := \operatorname{st}(x)$. Since f_N is S-continuous (by Exercise 4.25), we know that $f_N(x) \approx f_N(y)$. By Theorem 4.26, we know that f is (uniformly) continuous, whence $f(x) \approx f(y)$. Since y is standard and f_n converges pointwise to f, we know that $f_N(y) \approx f(y)$. Thus,

$$f_N(x) \approx f_N(y) \approx f(y) \approx f(x),$$

as desired. □

4.4. **Problems.**

Problem 4.1. Suppose $f : A \to \mathbb{R}$ and $c, L \in \mathbb{R}$.

(1) Show that $\lim_{x \to c^+} f(x) = L$ iff $f(x) \approx L$ for all $x \in A^*$ with $x \approx c$ and $x > c$.

(2) Show that $\lim_{x\to c^-} f(x) = L$ iff $f(x) \approx L$ for all $x \in A^*$ with $x \approx c$ and $x < c$.
(3) Show that $\lim_{x\to c} f(x) = L$ iff $f(x) \approx L$ for all $x \in A^*$ with $x \approx c$ and $x \neq c$.
(4) Show that $\lim_{x\to c^+} f(x) = L$ iff $f(x) \approx L$ for all $x \in A^*$ with $x \approx c$ and $x > c$.
(5) Show that $\lim_{x\to c} f(x) = +\infty$ iff $f(x) > \mathbb{N}$ for all $x \in A^*$ with $x \approx c$ and $x \neq c$.
(6) Show that $\lim_{x\to c} f(x) = -\infty$ iff $-f(x) > \mathbb{N}$ for all $x \in A^*$ with $x \approx c$ and $x \neq c$.
(7) Show that $\lim_{x\to+\infty} f(x) = L$ iff there is $x \in A^*$ such that $x > \mathbb{N}$ and $f(x) \approx L$ for all $x \in A^*$ with $x > \mathbb{N}$.
(8) Show that $\lim_{x\to-\infty} f(x) = L$ iff there is $x \in A^*$ such that $-x > \mathbb{N}$ and $f(x) \approx L$ for all $x \in A^*$ with $-x > \mathbb{N}$.

Problem 4.2. Suppose $f, g : A \to \mathbb{R}$ and $c, L, M \in \mathbb{R}$. Show that:
(1) $\lim_{x\to c} f(x) = L$ iff $\lim_{x\to c^+} f(x) = L$ and $\lim_{x\to c^-} f(x) = L$.
(2) If $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$, then:
  (a) $\lim_{x\to c}[(f+g)(x)] = L + M$;
  (b) $\lim_{x\to c}[(fg)(x)] = LM$;
  (c) $\lim_{x\to c}[\frac{f}{g}(x)] = \frac{L}{M}$ if $M \neq 0$.

Problem 4.3. Suppose that $\epsilon \in \mu$. Show that:
(1) $\sin(\epsilon) \approx 0$;
(2) $\cos(\epsilon) \approx 1$;
(3) $\tan(\epsilon) \approx 0$.

Problem 4.4. Determine the points of continuity for each of the functions below.
(1)
$$g(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$
(2)
$$h(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ -x & \text{if } x \notin \mathbb{Q}. \end{cases}$$
(3)
$$j(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ \frac{1}{n} & \text{if } x = \frac{m}{n} \in \mathbb{Q} \text{ in simplest form with } n \geq 1. \end{cases}$$

Problem 4.5. Suppose $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous and (s_n) is a Cauchy sequence. Show that $(f(s_n))$ is also a Cauchy sequence.

Problem 4.6. Suppose $f : [a, b] \to \mathbb{R}$ is monotonic on $[a, b]$. Further suppose that f satisfies the *Intermediate Value Property*, i.e. for any d in between $f(a)$ and $f(b)$, there is $c \in [a, b]$ such that $f(c) = d$. Show that f is continuous.

For $c \in \mathbb{R}$, define the *monad of* c to be $\mu(c) := \{x \in \mathbb{R}^* \mid x \approx c\}$.

Problem 4.7. There is a meta-principle, called *Cauchy's* principle, which states that if $c \in \mathbb{R}$ and a property holds for all x within some infinitesimal distance of c, then the property holds for all x within some real distance of c. Once we have the logical framework described precisely, we can give this a more definite meaning. This problem investigates some concrete instances of this principle.

(1) Suppose $c \in \mathbb{R}$ and $A \subseteq \mathbb{R}$. Show that $\mu(c) \subseteq A^*$ if and only if there is $\epsilon \in \mathbb{R}^{>0}$ such that $(c - \epsilon, c + \epsilon) \subseteq A$.
(2) Suppose $f : A \to \mathbb{R}$ and $\mu(c) \subseteq A^*$. Show that if f is constant on $\mu(c)$, then there is $\epsilon \in \mathbb{R}^{>0}$ such that $(c - \epsilon, c + \epsilon) \subseteq A$ and f is constant on $(c - \epsilon, c + \epsilon)$.
(3) Suppose $f : A \to \mathbb{R}$ is *continuous* and $f(x) \in \mathbb{R}$ for all $x \in A^*$ (i.e. the extension of f to A^* only takes standard values). Show that f is locally constant. (Hint: Use part (2).)

Problem 4.8. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *Lipshitz with constant* c if $c \in \mathbb{R}^{>0}$ and for any $x, y \in \mathbb{R}$, $|f(x) - f(y)| \le c|x - y|$. f is said to be a *contraction mapping* if f is Lipshitz with constant c for some $c < 1$. The *Contraction Mapping Theorem* says that a contraction mapping has a unique *fixed point*, that is, there is a unique $y \in \mathbb{R}$ such that $f(y) = y$. This theorem has many applications, for example in the proof of the existence and uniqueness of solutions of differential equations. In this exercise, we will prove the Contraction Mapping Theorem.

(1) Use nonstandard reasoning to show that a Lipshitz function is uniformly continuous.
(2) Show that a contraction mapping cannot have two distinct fixed points (this proves the uniqueness part of the Contraction Mapping Theorem).

Let f be a contraction mapping with Lipshitz constant $c \in (0, 1)$. Fix $x \in \mathbb{R}$. Recursively define a sequence (s_n) by $s_0 = x$ and $s_{n+1} = f(s_n)$. The goal is to show that (s_n) converges to a fixed point of f.

(3) Use induction to show that $|s_n - s_{n+1}| \le c^n |s_0 - s_1|$ for all $n \in \mathbb{N}$.
(4) Use the formula for the partial sums of a geometric series to show that $|s_0 - s_n| \le \frac{1-c^n}{1-c}|s_0 - s_1| \le \frac{1}{1-c}|s_0 - s_1|$ for all $n \in \mathbb{N}$.
(5) Use transfer of the result in (4) to show that $s_n \in \mathbb{R}_{\text{fin}}$ for $n > \mathbb{N}$. Fix $n > \mathbb{N}$ and set $y := \text{st}(s_n)$. Show that $s_n \approx s_{n+1}$. Conclude that $f(y) = y$.

It is interesting to remark that one can start with any $x \in \mathbb{R}$ and iterate f to approach the fixed point. The standard proof of the contraction mapping theorem involves proving the more careful estimate that

$$|s_m - s_n| \leq \frac{c^m}{1-c}|s_0 - s_1|$$

to show that (s_n) is Cauchy and hence converges.

5. Differentiation

5.1. **The derivative.** We suppose that $f : A \to \mathbb{R}$ is a function and $c \in A$ is an *interior point*, that is, there is an interval around c contained in A.

Definition 5.1. f is *differentiable at* c if $\lim_{h\to 0} \frac{f(c+h)-f(c)}{h}$ exists; in this case, the limit is denoted by $f'(c)$ and is called the *derivative of* f *at* c.

The nonstandard characterization of limits (Exercise 4.4) immediately gives the following nonstandard characterization of differentiability:

Proposition 5.2. *f is differentiable at c with derivative D if and only if for every $\epsilon \in \mu \setminus \{0\}$, we have $\frac{f(c+\epsilon)-f(c)}{\epsilon} \approx D$.*

Suppose f is differentiable at c. Fix $dx \in \mu \setminus \{0\}$. (Here we are calling our infinitesimal dx to match with the usual verbiage of calculus.) Set $df := df(c, dx) = f(c + dx) - f(c)$. Then $f'(x) = \operatorname{st}(\frac{df}{dx})$. In this way, the derivative is, in some sense, an actual fraction. (Recall in calculus we are warned not to take the notation $\frac{df}{dx}$ to seriously and not to treat this as an actual fraction.)

If the domain of f is an open (possibly infinite) interval, we say that f is *differentiable* if it is differentiable at all points of its domain.

Example 5.3. If $f(x) = x^2$, then

$$df = f(x + dx) - f(x) = (x + dx)^2 - x^2 = x^2 + 2xdx + (dx)^2 - x^2.$$

Thus, $\frac{df}{dx} = 2x + dx \approx 2x$, whence $f'(x)$ exists and $f'(x) = \operatorname{st}(\frac{df}{dx}) = 2x$.

Proposition 5.4. *If f is differentiable at x, then f is continuous at x.*

Proof. Suppose $y \approx x$; we need $f(y) \approx f(x)$. Write $y = x + dx$ with $dx \in \mu$. Without loss of generality, $dx \neq 0$. Then

$$f(y) - f(x) = df \approx f'(x)dx \approx 0.$$

□

Here are some fundamental properties of derivatives:

Theorem 5.5. *Suppose that f, g are differentiable at x and $c \in \mathbb{R}$. Then $f + g$, cf, and fg are also differentiable at x. If $g(x) \neq 0$, then $\frac{f}{g}$ is also differentiable at x. Moreover, the derivatives are given by:*

1. *$(f + g)'(x) = f'(x) + g'(x)$;*
2. *$(cf)'(x) = cf'(x)$;*
3. *(Product Rule) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$;*

(4) *(Quotient Rule)* $(\frac{f}{g})'(x) = \frac{f'(x)g(x)-f(x)g'(x)}{[g(x)]^2}$.

Proof. We'll only prove the Product Rule, leaving the others as exercises. Fix $dx \in \mu \setminus \{0\}$. Then

$$d(fg) = f(x+dx)g(x+dx) - f(x)g(x) = (f(x)+df)(g(x)+dg) - f(x)g(x),$$

so

$$\frac{d(fg)}{dx} = f(x)\frac{dg}{dx} + g(x)\frac{df}{dx} + (df)\frac{dg}{dx}.$$

Since f is continuous at x, $df \approx 0$. Thus, taking standard parts of the above display yields $(fg)'(x) = f(x)g'(x) + g(x)f'(x)$. □

Now, the Chain Rule is notorious for having many incorrect proofs in textbooks. Hopefully our nonstandard proof is correct:

Theorem 5.6 (Chain Rule)**.** *Suppose that f is differentiable at x and g is differentiable at $f(x)$. Then $g \circ f$ is differentiable at x and $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$.*

Proof. Fix $dx \in \mu \setminus \{0\}$. Then $df = f(x+dx) - f(x)$. Set $h := g \circ f$, so $dh = h(x+dx) - h(x)$. Then

$$dh = g(f(x+dx)) - g(f(x)) = g(f(x)+df) - g(f(x)).$$

Since f is continuous at x (by Proposition 5.4), $df \approx 0$. First suppose that $df \neq 0$. Then

$$\frac{dh}{dx} = \frac{dh}{df} \cdot \frac{df}{dx} = \frac{g(f(x)+df) - g(f(x))}{df} \cdot \frac{df}{dx} \approx g'(f(x)) \cdot f'(x).$$

Now suppose that $df = 0$. Then $f'(x) = 0$ and $dh = 0$, so $\frac{dh}{dx} = g'(f(x)) \cdot f'(x)$. Either way, $\frac{dh}{dx} \approx g'(f(x)) \cdot f'(x)$. Since dx was arbitrary, this shows that h is differentiable at x and $h'(x) = g'(f(x)) \cdot f'(x)$. □

The following theorem is the reason we use derivatives in connection with extrema of functions:

Theorem 5.7. *Suppose that $f : (a,b) \to \mathbb{R}$ is differentiable and f achieves* a local max *at $c \in (a,b)$, that is, there is $\epsilon \in \mathbb{R}^{>0}$ such that $f(x) \leq f(c)$ for all $x \in (c-\epsilon, c+\epsilon)$. Then $f'(c) = 0$.*

Proof. Suppose, towards a contradiction, that $f'(c) > 0$. (The case that $f'(c) < 0$ is similar.) Fix ϵ as in the statement of the theorem and fix $dx \in \mu^{>0}$. Then $c + dx \in (c-\epsilon, c+\epsilon)^*$, so $f(c+dx) \leq f(c)$. However, $f'(c) \approx \frac{f(c+dx)-f(c)}{dx}$, whence $\frac{f(c+dx)-f(c)}{dx} > 0$. Since $dx > 0$, this forces $f(c+dx) - f(c) > 0$, a contradiction. □

5.2. **Continuous differentiability.** In this subsection, $f : (a, b) \to \mathbb{R}$ is a function and $c \in (a, b)$. We aim to give a nonstandard characterization of f being *continuously differentiable*, that is f' exists and is continuous. First, let's give an alternate nonstandard characterization of differentiability.

Proposition 5.8. *f is differentiable at c if and only if, for all $u, u' \in (a, b)^*$ with $u \approx u' \approx c$ and $u \neq c$ and $u' \neq c$, we have*

$$\frac{f(u) - f(c)}{u - c} \approx \frac{f(u') - f(c)}{u' - c}.$$

Observe that this proposition has the desirable feature that it does not need to know the actual value of $f'(c)$ in order to verify differentiability; compare this with Proposition 5.2.

Proof. The ($\Rightarrow$) direction is immediate as both fractions are $\approx f'(c)$. We now prove the ($\Leftarrow$) direction. First, we show that f is continuous at c. Fix $u' \approx c$, $u' \neq c$. Set $L := |\frac{f(u')-f(c)}{u'-c}| \in \mathbb{R}^*$. Suppose $L \in \mathbb{R}_{\text{fin}}$. Then whenever $u \approx c$, $u \neq c$, since we have,

$$\frac{f(u) - f(c)}{u - c} - \frac{f(u') - f(c)}{u' - c} \approx 0,$$

it follows that $f(u) - f(c) - (\frac{f(u')-f(c)}{u'-c})(u - c) \approx 0$, and hence $f(u) \approx f(c)$. Now suppose that $L \notin \mathbb{R}_{\text{fin}}$. Fix $\epsilon \in \mu$, so $\frac{\epsilon}{L} \in \mu$. Suppose $|u - c| < \frac{\epsilon}{L}$. Then

$$|f(u) - f(c)| \leq |f(u) - f(c) - (\frac{f(u') - f(c)}{u' - c})(u - c)| + |(\frac{f(u') - f(c)}{u' - c})(u - c)|$$

The first term on the right hand side is infintiesimal and the second term is $\leq \epsilon$. Thus, if $|u - c| < \frac{\epsilon}{L}$, it follows that $f(u) \approx f(c)$. By the equivalence of (1) an (3) in Theorem 4.1, we see that f is continuous at c in this case too.

Now, fix $\delta \in \mathbb{R}^{>0}$ such that, whenever $u \in (a, b)$ and $|u - c| < \delta$, then $|f(u) - f(c)| < 1$. Suppose $u \in (a, b)^*$ and $|u - c| < \delta$. If $u \not\approx c$, then $\frac{f(u)-f(c)}{u-c} \in \mathbb{R}_{\text{fin}}$. If $u \approx c$, then $\frac{f(u)-f(c)}{u-c}$ might be infinite, but all such quotients (for different u's infinite close to c) are all infinitely close to one another. Thus:

$$\mathfrak{R}^* \models (\exists M \in \mathbb{R}^*)(\forall u \in (a, b)^*)(|u - c| < \delta \to |\frac{f(u) - f(c)}{u - c}| \leq M).$$

Applying transfer to this statement, we get a real bound $M \in \mathbb{R}^{>0}$ such that

$$\mathfrak{R} \models (\forall u \in (a, b))(|u - c| < \delta \to |\frac{f(u) - f(c)}{u - c}| \leq M).$$

But now, transfer this statement back to $\mathfrak{R}^*$: for all $u \in (a, b)^*$, if $|u-c| < \delta$, then $|\frac{f(u)-f(c)}{u-c}| \leq M$. In particular, for all $u \in (a, b)^*$, if $u \approx c$, then $\frac{f(u)-f(c)}{u-c} \in \mathbb{R}_{\text{fin}}$. The common standard part of all these fractions is then $f'(c)$. ☐

We can now give a similar nonstandard criteria for *continuous differentiability*:

Theorem 5.9. *$f : (a,b) \to \mathbb{R}$ is continuous differentiable if and only if, for all $u, u', v, v' \in (a,b)^*$ with $u \approx u' \approx v \approx v'$ and $u \neq v$ and $u' \neq v'$ and $\mathrm{st}(u) \in (a,b)$, we have*

$$\frac{f(u) - f(v)}{u - v} \approx \frac{f(u') - f(v')}{u' - v'}.$$

Proof. ($\Rightarrow$) Take $c \in (a,b)$ and fix $u, u', v, v' \approx c$ with $u \neq v$ and $u' \neq v'$. Then by the transfer of the Mean Value Theorem, there are w (resp. w') between u and v (resp. between u' and v') such that

$$f'(w) = \frac{f(u) - f(v)}{u - v} \quad \text{and } f'(w') = \frac{f(u') - f(v')}{u' - v'}.$$

Since $w, w' \approx c$ and f' is continuous at c, we have that $f'(w) \approx f'(w')$, yielding the desired result.

($\Leftarrow$) Fix $c \in (a,b)$. By considering $u, u' \approx c$, $u \neq c$, $u' \neq c$ and $v = v' = c$, the assumption and the previous proposition imply that $f'(c)$ exists. We now must show that f' is continuous at c. Fix $u \in (a,b)^*$, $u \approx c$; we need $f'(u) \approx f'(c)$. Fix $\epsilon \in \mu^{>0}$. By the transfer of the definition of f', there is $\delta \in (\mathbb{R}^*)^{>0}$ such that, for all $v \in (a,b)^*$, if $0 < |u - v| < \delta$, then $|\frac{f(u)-f(v)}{u-v} - f'(u)| < \epsilon$. Without loss of generality, we may assume that $\delta \in \mu$. Likewise, by the transfer of the fact that $f'(c)$ exists, we may further assume that δ is chosen small enough so that, for all $u' \in (a,b)^*$, if $0 < |u' - c| < \delta$, then $|\frac{f(u')-f(c)}{u'-c} - f'(c)| < \epsilon$. Choose $v, u' \in (a,b)^*$ such that $0 < |u - v|, |u' - c| < \delta$. Then:

$$|f'(u) - f'(c)| \leq \epsilon + |\frac{f(u) - f(v)}{u - v} - \frac{f(u') - f(c)}{u' - c}| + \epsilon.$$

By assumption, the middle term is infinitesimal, whence $f'(u) \approx f'(c)$. $\square$

5.3. **Problems.**

Problem 5.1. Suppose $f : (a,b) \to \mathbb{R}$ and $c \in (a,b)$. Show that f is differentiable at c with derivative D if and only if for every $\epsilon \in \mu \setminus \{0\}$, we have

$$\frac{f(c+\epsilon) - f(c)}{\epsilon} \approx D.$$

Problem 5.2. Suppose that $\epsilon \in \mu \setminus \{0\}$. Show that:

(1) $\frac{\sin(\epsilon)}{\epsilon} \approx 1$.
(2) $\frac{\cos(\epsilon)-1}{\epsilon} \approx 0$.
(3) $\frac{d}{dx}(\cos x) = -\sin x$.

Problem 5.3. Suppose that f and g are differentiable at x and $c \in \mathbb{R}$. Show that:

(1) cf is differentiable at x and $(cf)'(x) = cf'(x)$.
(2) If $g(x) \neq 0$, then $\frac{f}{g}$ is differentiable at x and $(\frac{f}{g})'(x) = \frac{g(x)f'(x)-f(x)g'(x)}{[g(x)]^2}$.

Problem 5.4. Suppose $f : \mathbb{R} \to \mathbb{R}$. We say that f has a *local maximum* at c if there is $\epsilon \in \mathbb{R}^{>0}$ such that $f(c) \geq f(x)$ for all $x \in (c - \epsilon, c + \epsilon)$. Show that f has a local maximum at c if and only if $f(c) \geq f(x)$ for all $x \approx c$. Define what it means for f to have a local minimum at c and give a similar nonstandard description.

Problem 5.5. Suppose $f : \mathbb{R} \to \mathbb{R}$ and $c \in \mathbb{R}$. Suppose that $f'(c) = 0$ and $f''(c) < 0$. Prove that f has a local maximum at c. Hint: Use transfer and the Mean Value Theorem to conclude that, for any $x \approx c$, $x \neq c$, there is t in between x and c such that

$$f'(t) = \frac{f(c) - f(x)}{c - x}.$$

(This result is often called the *Second Derivative Test.*)

The next few problems deal with the Taylor Series of a function. Suppose $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$. Further suppose that the n^{th} derivative of f at a exists for all n. Then the *Taylor series for f centered at a* is the power series

$$\sum_{i=0}^{\infty} \frac{f^{(i)}(x)}{i!}(x-a)^i = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots .$$

The Taylor series for f need not converge at some (or even any) x. Even if the Taylor series for f does converge at some x, it need not converge to $f(x)$. For example, suppose

$$f(x) = \begin{cases} e^{\frac{-1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then $f^{(n)}(x) = 0$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$, whence the Taylor series for f is identically 0.

Fix $n \geq 0$. The *n^{th} degree Taylor polynomial for f centered at a* is the polynomial

$$p_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(x)}{i!}(x-a)^i = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(x)}{n!}(x-a)^n.$$

Problem 5.6. For a given x, show that the Taylor series for f at x converges to f if and only if $p_n(x) \approx f(x)$ for all $n > \mathbb{N}$.

Set $R_n(x) := f(x) - p_n(x)$. It follows from the previous problem that the Taylor series for f at x converges to $f(x)$ if and only if $R_n(x) \approx 0$ for all $n > \mathbb{N}$. There is a theorem due to Lagrange that says that if f is $(n+1)$-times differentiable on some open interval I containing a, then for each $x \in I$, there is a c between a and x such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

Problem 5.7. Suppose that $f^{(n)}(x)$ exists for all $n \in \mathbb{N}$ and $x \in I$ (we say f is *infinitely differentiable on* I in this situation). Discuss how to make sense of $f^{(n)}(x)$ for $n \in \mathbb{N}^*$ and $x \in I^*$. Show that for all $x \in I^*$ and $n \in \mathbb{N}^*$ with $n \geq 1$, we have

$$p_n(x) - p_{n-1}(x) = \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Problem 5.8. Suppose that f is infinitely differentiable on I and $x \in I$.

(1) Suppose that
$$\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$
is infinitesimal for every $n > \mathbb{N}$ and every $c \in \mathbb{R}^*$ such that c is between a and x. Show that the Taylor series for f at x converges to $f(x)$.

(2) Use part (1) to show that the Taylor series for $\cos x$ centered at $a = 0$ (otherwise known as the *Maclaurin series for* $\cos x$) converges to $\cos x$ for all $x \in \mathbb{R}$.

(3) Use part (1) to show that the Maclaurin series for e^x converges to e^x for all $x \in \mathbb{R}$.

Problem 5.9. Suppose that $f^{(n)}$ exists for all real numbers in some open interval I. Further suppose that $f^{(n)}$ is continuous at $x \in I$. Show that for any infinitesimal Δx, there is an infinitesimal ϵ such that

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \frac{f''(x)}{2} + \cdots + \frac{f^{(n)}(x)}{n!}(\Delta x)^n + \epsilon(\Delta x)^n.$$

(Hint: Consider the Lagrange form of the remainder R_n.)

Problem 5.10. There is another form of the remainder R_n, which states that

$$R_n(x) = \frac{f^{(n)}(c) - f^{(n)}(a)}{(c-a)(n+1)!}(x-a)^{n+1}$$

for some c in between a and x. Apply this form of the remainder to R_{n-1} to prove the result from Problem 5.9 without using the assumption that $f^{(n)}$ is continuous.

6. Riemann Integration

6.1. **Hyperfinite Riemann sums and integrability.** The Riemann integral has a particularly slick description in terms of hyperfinite partitions. Indeed, one takes a Riemann sum with respect to rectangles of infinitesimal width; this Riemann sum will then be infinitely close to the Riemann integral.

First, let's recall some standard notions. Throughout, we assume that $f : [a,b] \to \mathbb{R}$ is a bounded function. A *partition of* $[a,b]$ is a finite ordered set $P = \{x_0, \ldots, x_n\}$ such that $a = x_0 < x_1 < \ldots < x_n = b$. A partition P_2 is a *refinement* of a partition P_1 if $P_1 \subseteq P_2$. (So P_2 is obtained from P_1 by further subdivision.) For a partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$, we define, for $i = 1, \ldots, n$:

- $M_i := M_i(P) := \sup\{f(x) \mid x \in [x_{i-1}, x_i]\}$
- $m_i := m_i(P) := \inf\{f(x) \mid x \in [x_{i-1}, x_i]\}$
- $\Delta x_i := x_i - x_{i-1}$.

Definition 6.1.

(1) The *upper Riemann sum of f with respect to P* is

$$U(f, P) := \sum_{i=1}^{n} M_i \Delta x_i.$$

(2) The *lower Riemann sum of f with respect to P* is

$$L(f, P) := \sum_{i=1}^{n} m_i \Delta x_i.$$

(3) The *left-hand Riemann sum of f with respect to P* is

$$S(f, P) := \sum_{i=1}^{n} f(x_{i-1}) \Delta x_i.$$

Set $M := \sup\{f(x) \mid x \in [a, b]\}$ and $m := \inf\{f(x) \mid x \in [a, b]\}$.

Exercise 6.2.

(1) For any partition P, we have

$$m(b - a) \leq L(f, P) \leq S(f, P) \leq U(f, P) \leq M(b - a).$$

(2) If P_2 is a refinement of P_1, then

$$L(f, P_1) \leq L(f, P_2) \leq U(f, P_2) \leq U(f, P_1).$$

(3) For any two partitions P_1 and P_2, $L(f, P_1) \leq U(f, P_2)$.

Part (2) of the previous exercise motivates the following

Definition 6.3.

(1) The *lower Riemann integral of f* is

$$L(f) := \sup\{L(f, P) \mid P \text{ a partition of } [a, b]\}.$$

(2) The *upper Riemann integral of f* is

$$U(f) := \inf\{L(f, P) \mid P \text{ a partition of } [a, b]\}.$$

By Exercise 6.2(1) and (3), we see that

$$m(b - a) \leq L(f) \leq U(f) \leq M(b - a).$$

We say that f is *Riemann integrable* if $U(f) = L(f)$. In this case, we set $\int_a^b f dx := U(f) = L(f)$.

The following Cauchy-type criterion for integrability is quite useful:

Exercise 6.4 (Riemann Lemma)**.** f is Riemann integrable if and only if, for every $\epsilon \in \mathbb{R}^{>0}$, there is a partition P of $[a,b]$ such that $U(f,P)-L(f,P) < \epsilon$.

Fix $\Delta x \in \mathbb{R}^{>0}$. Set $P_{\Delta x} = \{x_0, \ldots, x_n\}$, where $[a,x_1], [x_1,x_2], \ldots, [x_{n-2}, x_{n-1}]$ all have equal length Δx and $[x_{n-1}, x_n]$ is the "leftover" piece. (If $\Delta x \geq b-a$, then $P_{\Delta x} = \{a,b\}$.) We thus get a function $U(f,\cdot) : \mathbb{R}^{>0} \to \mathbb{R}$ by setting $U(f,\Delta x) := U(f, P_{\Delta x})$. In a similar manner, we get functions $L(f,\cdot)$ and $S(f,\cdot)$.

Here is another Cauchy-type criterion for integrability:

Exercise 6.5. f is Riemann integrable if and only if, for every $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $\Delta x \in \mathbb{R}^{>0}$, if $\Delta x < \delta$, then $U(f,\Delta x) - L(f,\Delta x)) < \epsilon$.

The functions $U(f,\cdot)$, $L(f,\cdot)$, $S(f,\cdot)$ have nonstandard extensions $U(f,\cdot) : (\mathbb{R}^*)^{>0} \to \mathbb{R}^*$, etc... So, for example, if $\Delta x \in \mu^{>0}$, then $U(f,\Delta x)$ equals the upper Riemann sum of f with respect to a hyperfinite partition, where each interval in the partition has infinitesimal length.

Theorem 6.6. *f is Riemann integrable if and only if $U(f,\Delta x) \approx L(f,\Delta x)$ for any $\Delta x \in \mu^{>0}$. In this case, for any $\Delta x \in \mu^{>0}$, we have*

$$\int_a^b f dx = \operatorname{st}(U(f,\Delta x)) = \operatorname{st}(L(f,\Delta x)) = \operatorname{st}(S(f,\Delta x)).$$

Proof. First suppose that f is Riemann integrable. Fix $\epsilon \in \mathbb{R}^{>0}$. Take $\delta \in \mathbb{R}^{>0}$ satisfying the conclusion of Exercise 6.5. By transfer, if $\Delta x \in \mu^{>0}$, then $U(f,\Delta x) - L(f,\Delta x) < \epsilon$. Since ϵ was arbitrary, this shows that $L(f,\Delta x) \approx U(f,\Delta x)$.

For the converse, we verify the criterion given by the Riemann Lemma. Let $\epsilon > 0$. By the assumption of the theorem,

$$\mathfrak{R}^* \models (\exists \Delta x \in (\mathbb{R}^*)^{>0})(U(f,\Delta x) - L(f,\Delta x) < \epsilon).$$

Now apply transfer.

It remains to verify the statement about the value of the integral. Fix $\Delta x \in \mu^{>0}$. Then $L(f,\Delta x) \leq S(f,\Delta x) \leq U(f,\Delta x)$ and $L(f,\Delta x) \approx U(f,\Delta x)$. Thus, $\operatorname{st}(U(f,\Delta x)) = \operatorname{st}(L(f,\Delta x)) = \operatorname{st}(S(f,\Delta x))$. Fix an ordinary partition P of $[a,b]$. Then, by the transfer of Exercise 6.2(3), we have

$$L(f,P) \leq U(f,\Delta x) \approx L(f,\Delta x) \leq U(f,P).$$

Thus, $L(f) \leq \operatorname{st}(U(f,\Delta x)) \leq U(f)$. Since f is Riemann integrable, we have

$$\int_a^b f dx = U(f) = L(f) = \operatorname{st}(U(f,\Delta x)).$$

□

We now verify that certain classes of functions are integrable.

Theorem 6.7. *If f is continuous on $[a,b]$, then for any $\Delta x \in \mu^{>0}$, $L(f,\Delta x) \approx U(f,\Delta x)$. Consequently, by the previous theorem, continuous functions are Riemann integrable.*

Proof. Let's give the idea of the proof first. Note that $U(f,\Delta x)-L(f,\Delta x) = \sum_{i=1}^n (M_i - m_i)\Delta x_i$. We will find an upper bound for this sum of the form $[f(c)-f(d)](b-a)$, where $|c-d| < \Delta x$. If $\Delta x \in \mu^{>0}$, then $c \approx d$, so $f(c) \approx f(d)$ by the uniform continuity of f. Since $b-a \in \mathbb{R}$, this will show that $U(f,\Delta x) - L(f,\Delta x) \approx 0$.

Now for the details: for $\Delta x \in \mathbb{R}^{>0}$, define $M_i(\Delta x) := M_i(P_{\Delta x})$ and $m_i(\Delta x) := m_i(P_{\Delta x})$. We then define the *oscillation of f with respect to Δx* to be the quantity

$$\omega(\Delta x) := \max\{M_i(\Delta x) - m_i(\Delta x) \mid i = 1, \ldots, n\}.$$

Suppose that $j \in \{1,\ldots,n\}$ is such that $\omega(\Delta x) = M_j(\Delta x) - m_j(\Delta x)$. Fix $c_{\Delta x}, d_{\Delta x} \in [x_{j-1}, x_j]$ such that $f(c_{\Delta x}) = M_j(\Delta x)$ and $f(d_{\Delta x}) = m_j(\Delta x)$; this is possible since continuous functions achieve their max and min. Clearly, $|c_{\Delta x} - d_{\Delta x}| \leq \Delta x$. Also,

$$U(f,\Delta x) - L(f,\Delta x) = \sum_{i=1}^n (M_i(\Delta x) - m_i(\Delta x))(\Delta x_i) \leq \omega(\Delta x)(b-a).$$

Fix $\Delta x \in \mu^{>0}$. By transfer, there are $c, d \in [a,b]^*$ such that $|c-d| \leq \Delta x$ and $U(f,\Delta x) - L(f,\Delta x) \leq (f(c)-f(d))(b-a)$, whence $U(f,\Delta x) \approx L(f,\Delta x)$. □

Exercise 6.8. Prove Theorem 6.7 with the assumption of "continuity" replaced by "monotonicity."

Now that integrals are infinitely close to hyperfinite sums, properties of integrals follow almost immediately from properties of sums. For example:

Proposition 6.9. *Suppose that f is integrable and $c \in \mathbb{R}$. Then cf is integrable and $\int_a^b (cf)dx = c\int_a^b f dx$.*

Proof. Fix $\Delta x \in \mu^{>0}$. First suppose that $c \geq 0$. Then $U(cf,\Delta x) = cU(f,\Delta x)$ and $L(cf,\Delta x) = cL(f,\Delta x)$. Since $U(f,\Delta x) \approx L(f,\Delta x)$, we have that $U(cf,\Delta x) \approx L(cf,\Delta x)$, whence cf is integrable. Moreover,

$$\int_a^b (cf)dx = \mathrm{st}(U(cf,\Delta x)) = c\,\mathrm{st}(U(f,\Delta x) = c\int_a^b f dx.$$

If $c < 0$, then $U(cf,\Delta x) = cL(f,\Delta x)$ and $L(cf,\Delta x) = cU(f,\Delta x)$. The proof then proceeds as in the previous paragraph. □

6.2. **The Peano Existence Theorem.** Here is an application of the nonstandard approach to integration to differential equations:

Theorem 6.10 (Peano Existence Theorem)**.** *Suppose that $g : [0,1] \times \mathbb{R} \to \mathbb{R}$ is a bounded, continuous function. Then for any $a \in \mathbb{R}$, there is a differentiable function $f : [0,1] \to \mathbb{R}$ satisfying $f(0) = a$ and $f'(t) = g(t, f(t))$ for all $t \in [0,1]$.*

Proof. The idea is quite simple: we find a "polygonal" solution by starting at $(0, a)$ and taking infinitesimal steps with slope according to g; we then take the standard part of the polygonal solution to obtain a curve solving the differential equation. Now for the gory details:

Let $Y : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ be such that, for all $0 \leq k < n$, we have $Y(0, n) = a$ and $Y(\frac{k+1}{n}, n) = Y(\frac{k}{n}, n) + g(\frac{k}{n}, Y(\frac{k}{n}, n)) \cdot \frac{1}{n}$. Let $Z : \mathbb{R} \to \mathbb{R}$ be such that $Z(\frac{k}{n}) = Y(\frac{k}{n}, n)$. Fix $N > \mathbb{N}$. We will show that $Z(\frac{l}{N}) \in \mathbb{R}_{\text{fin}}$ for all $0 \leq l \leq N$ and that $f : [0, 1] \to \mathbb{R}$ defined by $f(t) = \text{st}(Z(\frac{l}{N}))$, where $t \approx \frac{l}{N}$, is the desired solution to the differential equation.

Fix a bound $M \in \mathbb{R}^{>0}$ for g. Fix $0 \leq k \leq l \leq N$. Then:

$$\begin{aligned}
|Z(\frac{l}{N}) - Z(\frac{k}{N})| &= |\sum_{n=k}^{l-1}(g(\frac{n}{N}, Z(\frac{n}{N})) \cdot \frac{1}{N})| \\
&\leq \sum_{n=k}^{l-1}(|g(\frac{n}{N}, Z(\frac{n}{N}))| \cdot \frac{1}{N}) \\
&\leq M \cdot \sum_{n=k}^{l-1} \frac{1}{N} \\
&= M \cdot \frac{l-k}{N}.
\end{aligned}$$

This shows two things: first, by setting $k = 0$, we see that $|Z(\frac{l}{N})| \leq |a| + M \cdot \frac{l}{N}$ for all $0 \leq l \leq N$, whence $Z(\frac{l}{N}) \in \mathbb{R}_{\text{fin}}$. Secondly, if $\frac{l-k}{N} \approx 0$, then $Z(\frac{l}{N}) \approx Z(\frac{k}{N})$. These observations allow us to define $f : [0, 1] \to \mathbb{R}$ by setting $f(t) = \text{st}(Z(\frac{l}{N}))$ for any $l \in \{0, \ldots, N\}$ with $t \approx \frac{l}{N}$.

Exercise 6.11. Use the above calculation to verify that f is continuous.

It remains to verify that f is a solution of the differential equation $f'(t) = g(t, f(t))$. By the fundamental theorem of calculus, it is enough to show that $f(x) = a + \int_0^x g(t, f(t))dt$ for all $x \in [0, 1]$.

Let $W : \mathbb{R} \to \mathbb{R}$ be such that $W(\frac{k}{n}) = g(\frac{k}{n}, Z(\frac{k}{n}))$. Define $h : [0, 1] \to \mathbb{R}$ by $h(r) = g(r, f(r))$, a continuous function. Then, for $0 \leq k \leq N$, we have

$$W(\frac{k}{N}) \approx g(\text{st}(\frac{k}{N}), f(\text{st}(\frac{k}{N})) = h(\text{st}(\frac{k}{N})) \approx h(\frac{k}{N}).$$

For $n \in \mathbb{N}$, define $s_n := \max\{|W(\frac{k}{n}) - h(\frac{k}{n})| \mid 0 \leq k \leq n\}$. Thus, by the above observation, $s_N \in \mu$. Consequently, for $0 \leq k \leq N$, we have

$$\begin{aligned}
|\sum_{n=0}^{k}(W(\frac{n}{N}) \cdot \frac{1}{N}) - \sum_{n=0}^{k}(h(\frac{n}{N})\frac{1}{N})| &\leq \sum_{n=0}^{k} |W(\frac{n}{N}) - h(\frac{n}{N})| \cdot \frac{1}{N} \\
&\leq \sum_{n=0}^{k} s_N \cdot \frac{1}{N}
\end{aligned}$$

$$= s_N \cdot \frac{k+1}{N} \in \mu.$$

Therefore, for any $x \in [0,1]$, writing $x = \mathrm{st}(\frac{k}{N})$ with $\frac{k}{N} < x$, we have

$$\begin{aligned} f(x) &\approx Z(\frac{k}{N}) \\ &= a + \sum_{n=0}^{k-1}(g(\frac{n}{N}), Z(\frac{n}{N})) \cdot \frac{1}{N}) \\ &= a + \sum_{n=0}^{k-1}(W(\frac{n}{N}) \cdot \frac{1}{N}) \\ &\approx a + \sum_{n=0}^{k-1}(h(\frac{n}{N}) \cdot \frac{1}{N}) \\ &\approx a + \int_0^x h(t)dt. \end{aligned}$$

The last step follows from Theorem 6.6. Since the beginning and end are standard numbers, we have $f(x) = a + \int_0^x h(t)dt$. □

6.3. Problems.

Problem 6.1. Let $f : [a,b] \to \mathbb{R}$ be continuous. For $n \in \mathbb{N}^{>0}$ and $i = 0, \ldots, n-1$, define $x_i := a + \frac{i(b-a)}{n}$. Define

$$\mathrm{Av}(n) := \frac{f(x_0) + f(x_1) + \cdots + f(x_{n-1})}{n},$$

which is often referred to as a *sample average for* f. Prove that if $n > \mathbb{N}$, we have $\mathrm{Av}(n) \approx \frac{1}{b-a}\int_a^b f(x)dx$.

Remark. $\frac{1}{b-a}\int_a^b f(x)dx$ is often referred to as the *average of* f on $[a,b]$. This exercises illustrates a common phenomenon in nonstandard analysis, namely approximating continuous things by hyperfinite discrete things.

Problem 6.2 (Both)**.** Suppose that $f, g : [a,b] \to \mathbb{R}$ are Riemann integrable. Show that:

(1) $f+g$ is integrable and $\int_a^b (f(x)+g(x))dx = \int_a^b f(x)d(x) + \int_a^b g(x)dx$.
(2) f is Riemann integrable on both $[a,c]$ and $[c,b]$ for any $c \in [a,b]$ and $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_b^c f(x)dx$.
(3) $\int_a^b f(x)dx \leq \int_a^b g(x)dx$ if $f(x) \leq g(x)$ for all $x \in [a,b]$.
(4) $m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$ if $m \leq f(x) \leq M$ for all $x \in [a,b]$.

Problem 6.3. Suppose that $f : [a,b] \to \mathbb{R}$ is Riemann integrable. Define $F : [a,b] \to \mathbb{R}$ by $F(x) := \int_a^x f(t)dt$. Prove that F is continuous (even though f may not be).

Problem 6.4.

(1) Suppose $f : [a, \infty) \to \mathbb{R}$ is continuous. Define
$$\int_a^\infty f(x)dx := \lim_{b\to\infty} \int_a^b f(x)dx$$
if this limit exists; in this case the *improper integral* $\int_a^\infty f(x)dx$ is said to *converge*. Otherwise, the improper integral *diverges*. Show that $\int_a^\infty f(x)dx = L$ if and only if for every $b \in \mathbb{R}^*$ such that $b > \mathbb{N}$, we have $\int_a^b f(x)dx \approx L$. (But first: what does $\int_a^b f(x)dx$ for $b > \mathbb{N}$ even mean?)

(2) There are three other kinds of improper integrals, namely when f is continuous on intervals $(-\infty, a]$ as well as $[a, b)$ and $(a, b]$. Discuss how to define these improper integrals and give nonstandard characterizations of them as in (1).

Problem 6.5. Suppose $f : [a, \infty) \to \mathbb{R}$ is continuous and $f(x) \geq 0$ for all $x \in [a, \infty)$. Show that either $\int_a^\infty f(x)dx$ converges or diverges to $+\infty$. (Hint: Define $F : [a, \infty) \to \mathbb{R}$ by $F(x) := \int_a^x f(t)dt$. Consider the cases when F is bounded and unbounded respectively.)

Problem 6.6. Suppose that $(a_n : n \geq 1)$ is a sequence from $\mathbb{R}$. Suppose that $f : [1, \infty) \to \mathbb{R}$ is a continuous, nonnegative, nonincreasing function such that $f(n) = a_n$ for all $n \geq 1$. Show that $\sum_1^\infty a_n$ converges if and only if $\int_1^\infty f(x)dx$ converges. (This result is often called the *Integral Test*.)

7. Weekend Problem Set #1

Problem 7.1. (Sierpinski) Suppose $r, a_1, a_2, \ldots, a_n \in \mathbb{R}^{>0}$. Show that the equation
$$\frac{a_1}{x_1} + \frac{a_2}{x_2} + \cdots + \frac{a_n}{x_n} = r$$
has only finitely many solutions $(x_1, \ldots, x_n)$ in $\mathbb{N}^{>0}$.

Problem 7.2.

(1) Show there is $N \in \mathbb{N}^*$ such that N is divisible by k for every $k \in \mathbb{N}$.
(2) Let $P \subseteq \mathbb{N}$ denote the set of primes. Show that for every $n \in \mathbb{N}^*$, there is $p \in P^*$ such that p divides n.
(3) Use parts (1) and (2) to show that P is infinite (Euclid).
(4) Show that $\mathbb{Z}^*$ is a subring of $\mathbb{R}^*$.
(5) Let $p \in P^* \setminus P$. Let (p) be the ideal of $\mathbb{Z}^*$ generated by p. Show that the quotient ring $\mathbb{Z}^*/(p)$ is a field. What is the characteristic of $\mathbb{Z}^*/(p)$?

Problem 7.3. (Limit Comparison Test) Let $\sum_0^\infty a_i$ and $\sum_0^\infty b_i$ be two series, where $a_i, b_i > 0$ for all $i \in \mathbb{N}$. Suppose that $(\frac{a_i}{b_i})$ converges. Show that for $m, n > \mathbb{N}$ with $m \leq n$, we have $\sum_m^n a_i \in \mu$ if and only if $\sum_m^n b_i \in \mu$. Conclude that $\sum_0^\infty a_i$ converges if and only if $\sum_0^\infty b_i$ converges.

Problem 7.4. Let (s_n) denote a sequence in $\mathbb{R}$. For each $n \in \mathbb{N}$, define

$$\sigma_n := \frac{s_0 + \cdots s_n}{n+1}.$$

Show that if $s_n \to L$, then $\sigma_n \to L$. (Hint: If $N > \mathbb{N}$, there exists $M > \mathbb{N}$ such that $\frac{M}{N}$ is infinitesimal.)

We should say the sequence (σ_n) is called the sequence of *Cesáro means* of the sequence (s_n). It is possible that (σ_n) converges when (s_n) diverges. When (s_n) is the sequence of partial sums of an infinite series, this leads to the notion of *Cesáro summability*, which is useful in the theory of Fourier series.

Problem 7.5. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function and $f(x+y) = f(x)+f(y)$ for all $x, y \in \mathbb{R}$ (i.e. f is an additive group homomorphism).

(1) Show that $f(0) = 0$.
(2) Show that $f(nx) = nf(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$.
(3) Show that $f(-x) = -f(x)$ for all $x \in \mathbb{R}$.
(4) Show that $f(kx) = kf(x)$ for all $k \in \mathbb{Z}$.
(5) Show that $f(qx) = qf(x)$ for all $q \in \mathbb{Q}$.
(6) (Cauchy) Suppose f is *continuous*. Show that $f(x) = f(1) \cdot x$ for all $x \in \mathbb{R}$.

Problem 7.6. In this problem, we prove a strengthening of Cauchy's result from Problem 6 by showing that if $f : \mathbb{R} \to \mathbb{R}$ is an additive group homomorphism and there is an inteveral $I \subseteq \mathbb{R}$ such that f is bounded on I, then $f(x) = f(1) \cdot x$ for all $x \in \mathbb{R}$. (This result is due to Darboux.) Fix $x_0 \in I$ and $M \in \mathbb{R}^{>0}$ such that $|f(x)| \leq M$ for all $x \in I$.

(1) Show that if $x \approx 0$, then $|f(x)| \leq M + |f(x_0)|$.
(2) Show that if $x \approx 0$, then $f(x) \approx 0$. (Hint: If $x \approx 0$, then $nx \approx 0$ for all $n \in \mathbb{N}$.
(3) Show that $f(x) = f(1) \cdot x$ for all $x \in \mathbb{R}$. (Hint: Use the fact that any $x \in \mathbb{R}$ is infinitely close to an element of $\mathbb{Q}^*$.)

Problem 7.7.

(1) Suppose $f : (a, b) \to \mathbb{R}$ is C^1. Suppose $x \in (a, b)^*$ is such that $\mathrm{st}(x) \in (a, b)$. Suppose $\Delta x \approx 0$. Prove that there exists $\epsilon \in \mu$ such that
$$f(x + \Delta x) = f(x) + f'(x) + \epsilon\Delta x.$$
(2) Define $f : \mathbb{R} \to \mathbb{R}$ by
$$f(x) = \begin{cases} x^2 \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$
   (a) Show that $f'(x)$ exists for all x but that f' is not continuous at 0.

(b) Let $N > \mathbb{N}$ and $x = \frac{1}{2\pi N}$. Show that there is an infinitesimal Δx such that there is no $\epsilon \in \mu$ making the conclusion of (a) true.

(c) Discuss why parts (1) and (2)(b) don't contradict the fact that f' is continuous on $(0, 1)$.

Problem 7.8. A *Dirac delta function* is a definable function $D : \mathbb{R}^* \to \mathbb{R}^*$ such that:

- $D(x) \geq 0$ for all $x \in \mathbb{R}^*$;
- $\int_{-\infty}^{+\infty} D(x)dx = 1$;
- there is a positive infinitesimal $\delta > 0$ such that $\int_{-\delta}^{\delta} D(x)dx \approx 1$.

(1) Make sense of the above properties, i.e. explain how to precisely state the above properties of a Dirac delta function.

(2) Let D be a Dirac delta function and $f : \mathbb{R} \to \mathbb{R}$ a standard function. Show that $\mathrm{st}(\int_{-\infty}^{+\infty} f(x)D(x)dx) = f(0)$.

(3) Suppose $f : \mathbb{R} \to \mathbb{R}$ is standard and $\int_{-\infty}^{\infty} f(x)dx = 1$. Fix $n \in \mathbb{N}^* \setminus \mathbb{N}$ and define $D : \mathbb{R}^* \to \mathbb{R}^*$ by $D(x) := nf(nx)$. Show that D is a Dirac delta function. In particular, the following functions are Dirac delta functions:

- $$D(x) = \begin{cases} n & \text{if } |x| \leq \frac{1}{2n} \\ 0 & \text{otherwise.} \end{cases}$$
- $D(x) = \frac{n}{\pi(1+n^2x^2)}$.
- $D(x) = ne^{-\pi n^2 x^2}$.

(4) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a standard function such that $f(x) \geq 0$ for all $x \in \mathbb{R}$ and $\int_{-\infty}^{+\infty} f(x) = 1$. For $n \in \mathbb{N}^{>0}$, define $a_n := \int_{-\frac{1}{n}}^{\frac{1}{n}} f(x)$. Show that $a_n \approx 0$ for all $n > \mathbb{N}$. Conclude that a Dirac delta function can never be the nonstandard extension of a standard nonnegative, integrable function.

Problem 7.9. For $a, b \in \mathbb{N}$ with $a \leq b$, we set

$$[a, b] := \{a, a + 1, \ldots, b - 1, b\} \subseteq \mathbb{N}.$$

Suppose that $A \subseteq \mathbb{N}$. We say that:

- A is *thick* if for all $k \in \mathbb{N}^{>0}$, there is $x \in \mathbb{N}$ such that $[x+1, x+k] \subseteq A$.
- A is *syndetic* if $\mathbb{N} \setminus A$ is not thick, that is, there is $k \in \mathbb{N}^{>0}$ such that, for all $x \in \mathbb{N}$, $[x + 1, x + k] \cap A \neq \emptyset$.
- A is *piecewise syndetic* if $A = B \cap C$, where B is thick and C is syndetic.

(1) Prove that A is thick if and only if A^* contains an infinite interval, that is, there are $M, N \in \mathbb{N}^*$ with $N - M \in \mathbb{N}^* \setminus \mathbb{N}$ such that $[M, N] \subseteq A^*$.

(2) Prove that A is syndetic if and only if A^* has finite gaps, that is, for all intervals $[M, N] \subseteq \mathbb{N}^*$, if $[M, N] \cap A^* = \emptyset$, then $N - M \in \mathbb{N}$.
(3) (Standard) Prove that A is piecewise syndetic if and only if there is a finite set $F \subseteq \mathbb{N}$ such that $A + F$ is thick, where
$$A + F := \{a + f \ : \ a \in A, f \in F\}.$$
(4) Prove that A is piecewise syndetic if and only if there is an infinite interval on which A^* has only finite gaps.
(5) Use the nonstandard characterization of piecewise syndeticity to prove that piecewise syndeticity is a *partition regular* notion, that is, if $A = A_1 \cup \cdots \cup A_n$ is piecewise syndetic, then A_i is piecewise syndetic for some $i \in \{1, \ldots, n\}$.

The notions appearing in the previous problem are present in *additive combinatorics* and *combinatorial number theory.*

8. Many-sorted and Higher-Type Structures

We would now like to start applying nonstandard methods to areas of mathematics more complex than calculus. To do this, we will need a slightly more elaborate nonstandard framework.

8.1. **Many-sorted structures.**
In many areas of mathematics, we study many different sets at a time as well as functions between these various sets.

Example 8.1 (Linear Algebra)**.** A vector space is a set V together with two functions: vector addition, which is a function $+ : V \times V \to V$, and scalar multiplication, which is a function $\cdot : \mathbb{F} \times V \to V$, where $\mathbb{F}$ is some field.

Example 8.2 (Topology)**.** A metric space is a set X together with a metric, which is a function $d : X \times X \to \mathbb{R}$ satisfying certain axioms (to be defined in the next section).

Example 8.3 (Measure Theory)**.** A measure space is a triple $(X, \mathcal{B}, \mu)$, where X is a set, $\mathcal{B}$ is a σ-algebra of subsets of X (so $\mathcal{B} \subseteq \mathcal{P}(X)$), and $\mu : \mathcal{B} \to \mathbb{R}$ is a measure.

We now develop a nonstandard framework suitable for studying such situations. Before, we were working with a structure consisting of just a single "sort," namely a sort for $\mathbb{R}$. Now, we will work in a structure M with a (nonempty) collection of sorts $\mathcal{S}$. For each $s \in S$, we have a set M_s, the *universe of the sort s in M*. So, for example, in the linear algebra situation, we might have $\mathcal{S} = \{s, t\}$, with $M_s = V$ and $M_t = \mathbb{F}$. Often we will write a many-sorted structure as $M = (M_s \mid s \in \mathcal{S})$. Thus, we might write the linear algebra example as $(V, \mathbb{F})$, suppressing mention of the names of the sorts.

For any finite sequence $\vec{s} = (s_1, \ldots, s_n)$ of sorts, we have the product set $M_{\vec{s}} := M_{s_1} \times \cdots \times M_{s_n}$.

We now consider a language which was just as expressive as before. Namely, we have:

- For every finite sequence $\vec{s}$ of sorts and every $A \subseteq M_{\vec{s}}$, we have a predicate symbol P_A.
- For every sort s and every $a \in M_s$, we have a constant symbol c_a.
- For every finite sequence $\vec{s}$ of sorts, every sort t, and every function $f : M_{\vec{s}} \to M_t$, we have a function symbol F_f.

One now builds terms and formulae just as in ordinary logic, with the understanding that each sort comes equipped with its own collection of variables. If we need to be clear, we might decorate a variable with the name of the sort it is intended to range over, e.g. x^s.

Example 8.4. Returning to the vector space example, let's see how we might write the distributive law $c \cdot (x + y) = c \cdot x + c \cdot y$. Recall that $\mathcal{S} = \{s, t\}$, with $M_s = V$ and $M_t = \mathbb{F}$. Let $f : M_s \times M_s \to M_s$ denote vector addition and $g : M_t \times M_s \to M_s$ denote scalar multiplication. Then the axiom for the distributive law would be written as:

$$\forall x^s \forall y^s \forall z^t (g(z, f(x, y)) = f(g(z, x), g(z, y))).$$

Of course, for the purpose of sanity, in practice we will continue to write things as they might naturally be written in ordinary mathematics; however, one must be aware of the formal way that such sentences would be written.

As in the earlier part of these notes, we obtain a "nonstandard extension" by considering an embedding $M \to M^*$, where the universe of M^* corresponding to the sort s is M_s^* and the map is given by $a \mapsto a^*$. We will demand more from this embedding later, but for now we do not even assume that the embedding is an inclusion, that is, we do not assume that $M_s \subseteq M_s^*$. As before, we write A^* for the interpretation of the predicate symbol P_A in M^*.

To be of any use, we assume that the nonstandard extension is *proper*, meaning that, for every $s \in \mathcal{S}$ and every infinite $A \subseteq M_s$, we assume that there is $b \in A^*$ such that $b \neq a^*$ for any $a \in A$. (In other words, we are postulating the existence of many nonstandard elements.) Of course, such nonstandard extensions exist by the Compactness Theorem (Exercise!).

Exercise 8.5. Suppose the nonstandard extension is proper. Then for any finite sequence $\vec{s}$ of sorts and any infinite $A \subseteq M_{\vec{s}}$, A^* contains a nonstandard element.

8.2. **Higher-type sorts.** As we saw with the measure theory example in the previous subsection, it will often be convenient to have a sort for $\mathcal{P}(X)$ whenever X is itself a sort. For simplicity of discussion, let us consider the many-sorted structure $(X, \mathcal{P}(X))$; what we say now is easily adapted to the more general situation that X and $\mathcal{P}(X)$ are sorts in a many-sorted structure containing other sorts.

We have the nonstandard extension $(X, \mathcal{P}(X)) \to (X^*, \mathcal{P}(X)^*)$. We must be careful not to confuse $\mathcal{P}(X)^*$ with $\mathcal{P}(X^*)$, the latter retaining its usual meaning as the set of subsets of X^*. At the moment, $\mathcal{P}(X)^*$ is some abstract set, perhaps having no affiliation with an actual powerset. We now discuss how to relate $\mathcal{P}(X)^*$ and $\mathcal{P}(X^*)$.

Set $E = \{(x, A) \in X \times \mathcal{P}(X) \mid x \in A\}$, the symbol for the *membership relation.*

Lemma 8.6 (Normalization). *We may assume that our nonstandard extension satisfies the additional two conditions:*

(N1) $X \subseteq X^*$ *and* $x = x^*$ *for all* $x \in X$*;*

(N2) $\mathcal{P}(X)^* \subseteq \mathcal{P}(X^*)$ *and* E^* *is the membership relation restricted to* $X^* \times \mathcal{P}(X)^*$.

Proof. We begin with some abstract nonsense: let Y be a set and $h : X^* \to Y$ a bijection such that $X \subseteq Y$ and $h(x^*) = x$ for all $x \in X$.

For $A \in \mathcal{P}(X)^*$, set $\Phi(A) = \{h(x) \mid x \in X^* \text{ and } (x, A) \in E^*\} \subseteq Y$. We claim that Φ is injective. Indeed, suppose that $A_1, A_2 \in \mathcal{P}(X)^*$ and $A_1 \neq A_2$. By the transfer principle, there is $x \in X^*$ such that either $(x, A_1) \in E^*$ and $(x, A_2) \notin E^*$; or $(x, A_2) \in E^*$ and $(x, A_1) \notin E^*$. Then either $h(x) \in \Phi(A_1) \setminus \Phi(A_2)$ or $h(x) \in \Phi(A_2) \setminus \Phi(A_1)$; either way, $\Phi(A_1) \neq \Phi(A_2)$.

Now make $(Y, \Phi(\mathcal{P}(X)^*))$ into a structure in the unique way so that the map $(h, \Phi) : (X^*, \mathcal{P}(X)^*) \to (Y, \Phi(\mathcal{P}(X)^*))$ is an isomorphism. (Exercise!) Note that $(Y, \Phi(\mathcal{P}(X)^*))$ has the desired properties (N1) and (N2). □

Recap: We have the many-sorted structure $(X, \mathcal{P}(X))$ and its nonstandard extension $(X^*, \mathcal{P}(X)^*)$. Furthermore, we assume that $X \subseteq X^*$ and for all $Y \in \mathcal{P}(X)^*$, we view $Y \subseteq X^*$ by declaring, for $x \in X^*$:

$$x \in Y \leftrightarrow (x, Y) \in E^*.$$

There is some potential confusion that we should clear up now. Suppose that $A \subseteq X$. Then we have $A^* \subseteq X^*$ from the interpretation of the symbol P_A. However, $A \in \mathcal{P}(X)$, so it is mapped by the embedding to an element of $\mathcal{P}(X)^*$, which we temporarily denote by $(A)^*$. Fortunately, all is well:

Lemma 8.7. $A^* = (A)^*$.

Proof. By the transfer principle, we have that, for $x \in X^*$, $x \in A^*$ if and only if $(X^*, \mathcal{P}(X)^*) \models P_A(x)$. By the normalization assumption, we have that, for $x \in X^*$, $x \in (A)^*$ if and only if $(x, (A)^*) \in E^*$. Fortunately, $(X, \mathcal{P}(X)) \models \forall x \in X(P_A(x) \leftrightarrow P_E(x, A))$, so the desired result follows from transfer. □

Definition 8.8. A subset A of X^* is called *internal* if $A \in \mathcal{P}(X)^*$; otherwise, A is called *external.*

Thus, the transfer principle applies to the internal subsets of X^*.

Example 8.9. Let us consider $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ and its nonstandard extension $(\mathbb{N}^*, \mathcal{P}(\mathbb{N})^*)$. We claim that $\mathbb{N}$ is an *external* subset of $\mathbb{N}^*$. To see this, note that the following sentence is true in $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$:

$$\forall A \in \mathcal{P}(\mathbb{N})((\exists x \in \mathbb{N}(P_E(x,A)) \wedge \exists y \in \mathbb{N} \forall z \in \mathbb{N}(P_E(z,A) \to z \leq y))$$
$$\to \exists y \in \mathbb{N}(P_E(y,A) \wedge \forall z \in \mathbb{N}(P_E(z,A) \to z \leq y).$$

This sentence says that if $A \subseteq \mathbb{N}$ is bounded above, then A has a maximum element. By transfer, the same holds true for any $A \in \mathcal{P}(\mathbb{N})^*$, that is, for any *internal* subset of $\mathbb{N}^*$. If $\mathbb{N}$ were internal, then since it is bounded above (by an infinite element), it would have a maximum, which is clearly not true.

Example 8.10. We continue to work with the set-up of the previous example. Since

$$(\mathbb{N}, \mathcal{P}(\mathbb{N})) \models \forall n \in \mathbb{N} \exists A \in \mathcal{P}(\mathbb{N}) \forall m \in \mathbb{N}(P_E(m,A) \leftrightarrow m \leq n),$$

by transfer we have

$$(\mathbb{N}^*, \mathcal{P}(\mathbb{N})^*) \models \forall n \in \mathbb{N}^* \exists A \in \mathcal{P}(\mathbb{N})^* \forall m \in \mathbb{N}^*(P_E(m,A) \leftrightarrow m \leq n).$$

Fixing $N \in \mathbb{N}$, we suggestively let $\{0, 1, \ldots, N\}$ denote the internal subset of $\mathbb{N}^*$ consisting of all the elements of $\mathbb{N}^*$ that are no greater than N. This is a prototypical example of a *hyperfinite set*.

The following principle is useful in practice; it says that sets defined (in the first-order logic sense) from internal parameters are internal.

Theorem 8.11 (Internal Definition Principle)**.** *Let $\varphi(x, x_1, \ldots, x_m, y_1, \ldots, y_n)$ be a formula, where $x, x_1, \ldots, x_m$ range over the sort for X and $y_1, \ldots, y_m$ range over the sort for $\mathcal{P}(X)$. Suppose that $a_1, \ldots, a_m \in X^*$ and $A_1, \ldots, A_n \in \mathcal{P}(X)^*$. Set*

$$B := \{b \in X^* \mid (X^*, \mathcal{P}(X)^*) \models \varphi(b, a_1, \ldots, a_m, A_1, \ldots, A_n)\}.$$

Then B is internal.

Proof. The following sentence is true in $(X, \mathcal{P}(X))$:

$$\forall x_1, \ldots, x_m \forall y_1, \ldots, y_n \exists z \forall x(P_E(x,z) \leftrightarrow \varphi(x, x_1, \ldots, x_m, y_1, \ldots, y_n)).$$

By transfer, this remains true in $(X^*, \mathcal{P}(X)^*)$. Plugging in a_i for x_i and A_j for y_j, we see that

$$(X^*, \mathcal{P}(X)^*) \models \exists z \forall x(P_E(x,z) \leftrightarrow \varphi(x, a_1, \ldots, a_m, A_1, \ldots, A_n)).$$

The set asserted to exist is B, which then belongs to $\mathcal{P}(X)^*$, that is, B is internal. □

Example 8.12. For any finite collection $a_1, \ldots, a_m \in X^*$, the set $\{a_1, \ldots, a_m\}$ is internal. Indeed, let $\varphi(x, x_1, \ldots, x_m)$ be the formula $x = x_1 \vee \cdots \vee x = x_m$. Then

$$\{a_1, \ldots, a_m\} = \{b \in X^* \mid (X^*, \mathcal{P}(X)^*) \models \varphi(b, a_1, \ldots, a_m)\}.$$

It will also prove useful to have a notion of internal function. To do this, we need to expand our set-up a bit. We now consider the many-sorted structure $(X, \mathcal{P}(X), \mathcal{P}(X \times X))$ with an embedding into a nonstandard extension $(X^*, \mathcal{P}(X)^*, \mathcal{P}(X \times X)^*)$. We set:

- $E_1 := \{(x, A) \in X \times \mathcal{P}(X) \mid x \in A\}$, and
- $E_2 := \{(x, y, A) \in X \times X \times \mathcal{P}(X \times X) \mid (x, y) \in A\}$.

The proof of the following lemma is exactly like the proof of Lemma 8.6.

Lemma 8.13. *We may assume that our nonstandard extension satisfies the additional three conditions:*

(N1) $X \subseteq X^*$ *and* $x = x^*$ *for all* $x \in X$*;*
(N2) $\mathcal{P}(X)^* \subseteq \mathcal{P}(X^*)$ *and* E_1^* *is the membership relation restricted to* $X^* \times \mathcal{P}(X)^*$*;*
(N3) $\mathcal{P}(X \times X)^* \subseteq \mathcal{P}(X^* \times X^*)$ *and* E_2^* *is the membership relation restricted to* $X^* \times X^* \times \mathcal{P}(X \times X)^*$*.*

Definition 8.14. $B \subseteq X^* \times X^*$ is *internal* if $B \in \mathcal{P}(X \times X)^*$. If $A, B \subseteq X^*$ and $f : A \to B$ is a function, then we say that f is *internal* if the graph of f, $\Gamma(f) := \{(x, y) \in X^* \times X^* \mid x \in A \text{ and } f(x) = y\} \subseteq X^* \times X^*$, is internal.

At this point, the reader should verify that they would know how to escape the friendly confines of considering many-sorted structures of the form $(X, \mathcal{P}(X))$ or $(X, \mathcal{P}(X), \mathcal{P}(X \times X))$ and instead be able to consider much wilder many-sorted structures that might contain many sets and their powersets. For example, in studying vector spaces, it will be convenient to consider a many-sorted structure of the form $(V, \mathcal{P}(V), \mathcal{P}(V \times V), \mathbb{F}, \mathcal{P}(\mathbb{F}), \mathcal{P}(\mathbb{F} \times V))$. As an exercise, make sure you understand how to speak of an internal linear transformation $T : V^* \to V^*$ or an internal norm $\| \cdot \| : V^* \to \mathbb{R}^*$.

In a similar vein, if $s : \mathbb{N}^* \to X^*$ is an internal function, then we refer to the "sequence" $(s_n \mid n \in \mathbb{N}^*)$ as an *internal sequence.*

Definition 8.15. Set $\mathcal{P}_{\text{fin}}(X) := \{A \in \mathcal{P}(X) \mid A \text{ is finite }\} \subseteq \mathcal{P}(X)$. We then say that $B \subseteq X^*$ is *hyperfinite* if $B \in \mathcal{P}_{\text{fin}}(X)^*$.

Exercise 8.16. Assume that $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\mathbb{N} \times \mathbb{N}))$ is part of the structure.

(1) Prove that hyperfinite sets are internal.
(2) Prove that an internal subset of a hyperfinite set is hyperfinite.
(3) Prove that $B \subseteq X^*$ is hyperfinite if and only if there is an internal function $f : B \to \mathbb{N}^*$ such that f is a bijection between B and $\{0, 1, \ldots, N\}$ for some $N \in \mathbb{N}^*$; we then refer to $N+1$ as the *internal cardinality of* B.
(4) Prove that finite subsets of X^* are hyperfinite.

8.3. **Saturation.** For various nonstandard arguments, it will not suffice to merely assume that the extension is proper; we will need to assume a further *richness* condition on our nonstandard extensions.

Definition 8.17. Suppose that κ is an uncountable cardinal. We say that the nonstandard extension is *κ-saturated* if whenever $(A_i \mid i \in I)$ is a family of internal sets with $|I| < \kappa$ satisfying the *finite intersection property*, that is, the intersection of any finite number of A_i's is nonempty, then $\bigcap_{i\in I} A_i \neq \emptyset$.

Example 8.18. Suppose that the nonstandard extension is $\aleph_1$-saturated and that $\mathbb{R}$ is a basic sort. For each $q \in \mathbb{Q}^{>0}$, set $A_q := \{r \in \mathbb{R}^* \mid 0 < r < q\}$. By the internal definition principle, A_q is an internal set. Moreover, it is easy to verify that the family $(A_q \mid q \in \mathbb{Q}^{>0})$ has the finite intersection property. By $\aleph_1$-saturation, there is $r \in \bigcap_{q\in\mathbb{Q}^{>0}} A_q$; this r is then a positive infinitesimal.

Although this example is modest (as we already have ways of constructing infinitesimals), saturation will prove crucial in the analysis to come; see also the problems below.

Exercise 8.19. Assume that the nonstandard extension is $\aleph_1$-saturated and that $(N_m \mid m \in \mathbb{N})$ is a sequence of elements of $\mathbb{N}^* \setminus \mathbb{N}$. Prove that there is $N > \mathbb{N}$ such that $N < N_m$ for each $m \in \mathbb{N}$.

In any research article, one always assumes at least $\aleph_1$-saturation of the nonstandard extension. For other applications, κ-saturation for larger κ is often needed. However, how can we be assured that κ-saturated nonstandard extensions exist?

Theorem 8.20. *For any uncountable cardinal κ, there is a κ-saturated nonstandard extension.*

Proof. See any textbook on model theory. □

Saturation is closely related to another richness concept, namely *comprehension.*

Theorem 8.21 (Saturated extensions are comprehensive)**.** *Suppose that the nonstandard extension is κ-saturated. Suppose A and B are internal sets and suppose that $A_0 \subseteq A$ is a (possibly external) set with $|A_0| < \kappa$. Suppose $f_0 : A_0 \to B$ is a function. Then there is an* internal *function $f : A \to B$ extending f_0.*

Proof. For $x \in A_0$, define

$$D_x := \{f \mid f : A \to B \text{ is internal and } f(x) = f_0(x)\}.$$

By the internal definition principle, each D_x is internal. By κ-saturation, it remains to verify that $(D_x \mid x \in A_0)$ has the finite intersection property. Fix $x_1, \ldots, x_n \in A_0$. Define $f : A \to B$ by

$$f(y) = \begin{cases} f_0(x_i) & \text{if } y = x_i \text{ for some } i \in \{1, \ldots, n\}; \\ f_0(x_1) & \text{otherwise.} \end{cases}$$

Then f is internal by the internal definition principle. □

Definition 8.22. The nonstandard extension is said to be *countably comprehensive* if whenever B is internal and $(b_n \mid n \in \mathbb{N})$ is a countable sequence from B, then there is an internal $f : \mathbb{N}^* \to B$ such that $f(n) = b_n$ for all $n \in \mathbb{N}$.

In other words, countable comprehension says that sequences indexed by $\mathbb{N}$ can be *internally extended* to sequences indexed by $\mathbb{N}^*$.

Corollary 8.23. *If the nonstandard extension is $\aleph_1$-saturated, then it is also countably comprehensive.*

The converse to the previous corollary is also true; see Problem 8.8 below.

8.4. **Useful nonstandard principles.** In this subsection, we collect some useful principles that are often used in nonstandard proofs.

Theorem 8.24 (Overflow)**.** *Suppose that $A \subseteq \mathbb{R}^*$ is internal and suppose that, for every $n \in \mathbb{N}$, there is $a \in A$ such that $a > n$. Then there is $a \in A$ such that $a > \mathbb{N}$.*

Proof. Set $B := \{n \in \mathbb{N}^* \mid \text{ there is } a \in A \text{ such that } a > n\}$. B is internal by the internal definition principle. By assumption, $\mathbb{N} \subseteq B$. Since $\mathbb{N}$ is external, we have $\mathbb{N} \subsetneq B$, whence there is $n \in B$ with $n > \mathbb{N}$. Take $a \in A$ such that $a > n$; it follows that $a > \mathbb{N}$. $\square$

Theorem 8.25 (Underflow)**.** *Let $A \subseteq (\mathbb{R}^*)^{>0}$ be internal and suppose that, for every $n \in \mathbb{N}^* \setminus \mathbb{N}$, there is $a \in A$ such that $a < n$. Then $A \cap \mathbb{R}_{\text{fin}} \neq \emptyset$.*

Proof. Set $B := \{n \in \mathbb{N}^* \mid \text{ there is } a \in A \text{ such that } a < n\}$. Then B is internal and contains $\mathbb{N}^* \setminus \mathbb{N}$. Note that $\mathbb{N}^* \setminus \mathbb{N}$ is external, else, since $\mathbb{N}^*$ is internal, we would see that $\mathbb{N}$ is internal. Thus, we have $B \cap \mathbb{N} \neq \emptyset$, which yields the desired result. $\square$

Theorem 8.26 (Internal Induction)**.** *Let $A \subseteq \mathbb{N}^*$ be internal and suppose that:*

- *$0 \in A$;*
- *for all $n \in \mathbb{N}^*$, if $n \in A$, then $n + 1 \in A$.*

Then $A = \mathbb{N}^*$.

Proof. Suppose $A \neq \mathbb{N}^*$. Since $\mathbb{N}^* \setminus A$ is internal, it has a minimum element (by transfer). Let $n = \min(\mathbb{N}^* \setminus A)$. By assumption, $n > 0$. Then $n - 1 \in A$, whence, by assumption, $n \in A$, a contradiction. $\square$

As a warning, we really must assume that A is internal in the previous theorem. Indeed, $\mathbb{N}$ satisfies the two assumptions of the previous theorem, but $\mathbb{N} \neq \mathbb{N}^*$.

Theorem 8.27 (Infinitesimal Prolongation)**.** *Suppose that $(s_n \mid n \in \mathbb{N}^*)$ is an internal sequence from $\mathbb{R}^*$. Suppose that $s_n \approx 0$ for each $n \in \mathbb{N}$. Then there is $N > \mathbb{N}$ such that $s_n \approx 0$ for each $n \leq N$.*

Incorrect First Attempt. Let $A = \{n \in \mathbb{N}^* \mid s_n \not\approx 0\}$. If $A = \emptyset$, we are done. Otherwise, let $N = \min A$. By assumption, $N > \mathbb{N}$, whence we are done. This proof is incorrect since A is *external*, whence we cannot conclude that it has a minimum! □

Proof. For $n \in \mathbb{N}^*$, define $t_n := n \cdot s_n$; then the sequence $(t_n \mid n \in \mathbb{N}^*)$ is internal by assumption. By assumption, $t_n \approx 0$ for $n \in \mathbb{N}$; again, this is an external statement. However, it is enough to replace that statement by the weaker, *internal* statement $|t_n| < 1$ for all $n \in \mathbb{N}$. More precisely, let $A = \{n \in \mathbb{N}^* \mid |t_m| < 1 \text{ for all } m \leq n\}$. By overflow, there is $N \in \mathbb{N}^* \setminus \mathbb{N}$ such that $N \in A$. This N is as desired. □

The Infinitesimal Prolongation Theorem is often used in conjunction with countable comprehension as follows. Suppose that $(x_n \mid n \in \mathbb{N}^*)$ is an internal sequence from $\mathbb{R}^*$ such that $x_n \in \mathbb{R}_{\text{fin}}$ for $n \in \mathbb{N}$. For $n \in \mathbb{N}$, define $y_n := \text{st}(x_n)$. By countable comprehension, we can internally extend the sequence $(y_n \mid n \in \mathbb{N})$ to an internal sequence $(y_n \mid n \in \mathbb{N}^*)$ from $\mathbb{R}^*$. For $n \in \mathbb{N}$, we know that $x_n \approx y_n$. By infinitesimal prolongation, we can find $N > \mathbb{N}$ such that $x_n \approx y_n$ for all $n \leq N$.

The last principle is perhaps the most important of all in applications, as we will see later in these notes.

Theorem 8.28 (Hyperfinite Approximation)**.** *Suppose that the nonstandard extension is κ-saturated. Let $A \subseteq X^*$ be internal and let $B \subseteq A$ be a (possibly external) set with $|B| < \kappa$. Then there is a hyperfinite set $C \subseteq A$ such that $B \subseteq C$.*

Proof. For $x \in B$, set

$$D_x := \{C \in \mathcal{P}_{\text{fin}}(X)^* \mid C \subseteq A \text{ and } x \in C\}.$$

Note that each D_x is internal by the internal definition principle. Moreover, $(D_x \mid x \in B)$ has the finite intersection property. Indeed, given $x_1, \ldots, x_n \in B$, we have that $\{x_1, \ldots, x_n\} \in D_{x_1} \cap \cdots \cap D_{x_n}$. Thus, by κ-saturation, there is $C \in \bigcap_{x \in B} D_x$; such C is as desired. □

8.5. **Recap: the nonstandard setting.** From now on, we will proceed under the following assumptions: we assume that our nonstandard extension contains as sorts all sets relevant to the mathematics we are about to study. Moreover, we assume that enough cartesian products and powersets are also sorts. We assume that the nonstandard extension satisfies the normalization assumptions from Lemma 8.6. Finally, we assume that the nonstandard extension is κ-saturated for κ large enough for our purposes.

8.6. **Problems.**

Problem 8.1. Suppose that A and B are internal subsets of X^*.

(1) Show that $X^* \setminus A$, $A \cup B$ and $A \cap B$ are internal.
(2) Suppose that $f : X^* \to X^*$ is an internal function. Show that $f(A)$ and $f^{-1}(A)$ are internal sets and $f \restriction A$ is an internal function.

Problem 8.2.

(1) Suppose $r, s \in \mathbb{R}^*$ and $r < s$. Set $[r, s] := \{t \in \mathbb{R}^* \mid r \leq t \leq s\}$. Show that $[r, s]$ is internal.
(2) Show that μ is external.
(3) Show that $\mathbb{R}_{\text{fin}}$ is external.

Problem 8.3. Discuss what it should mean for a function $f : A \to B$ to be internal, where $A \subseteq M_{\vec{s}}$ and $B \subseteq M_{\vec{t}}$.

Problem 8.4. Suppose that $f : \mathbb{N}^* \times X^* \to X^*$ is an internal function. Fix $x \in X^*$. Show that there exists a unique internal function $F : \mathbb{N}^* \to X^*$ such that $F(0) = x$ and $F(n+1) = f(n+1, F(n))$. (This is the principle of Internal Recursion.)

Problem 8.5. Suppose that the nonstandard extension is κ-saturated. Show that every infinite internal set has cardinality at least κ.

Problem 8.6.

(1) Suppose that $N \in \mathbb{N}^* \setminus \mathbb{N}$. Fix $r \in (0, 1)$ (so r is standard). Show there is a smallest $k \in \mathbb{N}^*$ such that $Nr \leq k$.
(2) Show that any infinite hyperfinite set has cardinality at least $2^{\aleph_0}$.
(3) Show that any infinite internal set has cardinality at least $2^{\aleph_0}$. (This improves the result from the previous exercise.)

Problem 8.7. Suppose the nonstandard extension satisfies the Countable Comprehension Principle. Further suppose that $(K_n \mid n \in \mathbb{N})$ is a sequence of elements of $\mathbb{N}^*$ such that $K_n > \mathbb{N}$ for all $n \in \mathbb{N}$. Show that there is $K \in \mathbb{N}^* \setminus \mathbb{N}$ such that $K < K_n$ for all $n \in \mathbb{N}$.

Problem 8.8. Show that a nonstandard extension satisfying the Countable Comprehension Principle must be $\aleph_1$-saturated. (Hint: you might find the previous problem useful.)

Problem 8.9. Fix $k \in \mathbb{N}$. Suppose that $G = (V, E)$ is a (combinatorial) graph such that every finite subgraph of G is k-colorable. Prove that G is k-colorable. (Hint: hyperfinite approximation!)

9. Metric Space Topology

In this section, we will start using the new nonstandard framework developed in the previous section to develop metric space topology from a nonstandard point of view. Although general topological spaces can be treated in the nonstandard framework, we have made a conscience decision to discuss the (important) special case of metric spaces.

9.1. Open and closed sets, compactness, completeness.

Definition 9.1. A *metric space* is a pair (X, d) such that X is a nonempty set and $d : X \times X \to \mathbb{R}$ is a function (the *metric* or *distance function*) satisfying, for all $x, y, z \in X$:

(1) $d(x,y) \geq 0$;
(2) $d(x,y) = 0$ if and only if $x = y$;
(3) $d(x,y) = d(y,x)$;
(4) (Triangle Inequality) $d(x,z) \leq d(x,y) + d(y,z)$.

If (2) in the above list is replaced by the weaker condition

(2') $x = y$ implies $d(x,y) = 0$,

then d is called a *pseudometric* and (X,d) is called a *pseudometric space*.

As usual, we often speak of "the metric space X," suppressing mention of the metric. Keep in mind: the same set X can be equipped with many different metrics, yielding many different metric spaces. (See the examples below.)

Example 9.2.

(1) For $\vec{x}, \vec{y} \in \mathbb{R}^n$, define $d(\vec{x}, \vec{y}) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$. Then $(\mathbb{R}^n, d)$ is a metric space. This metric is usually referred to as the *euclidean metric* on $\mathbb{R}^n$.
(2) For $\vec{x}, \vec{y} \in \mathbb{R}^n$, define $d_\infty(\vec{x}, \vec{y}) := \max_{i=1,\ldots,n} |x_i - y_i|$. Then $(\mathbb{R}^n, d_\infty)$ is a metric space.
(3) Set $C([0,1]), \mathbb{R})$ to be the set of continuous functions $f : [0,1] \to \mathbb{R}$. Define a metric d on $C([0,1], \mathbb{R})$ by setting
$$d(f,g) := \max_{x \in [0,1]} |f(x) - g(x)|.$$
Note that this maximum exists by continuity.

Until otherwise specified, we fix a metric space X. For $a \in X$ and $r \in \mathbb{R}^{>0}$, we set $B(a;r) := \{x \in X \mid d(a,x) < r\}$, the *open ball in X centered at a with radius r*. We can also consider the *closed ball in X centered at a with radius r*: $\bar{B}(a;r) := \{x \in X \mid d(a,x) \leq r\}$.

We will work in a nonstandard extension containing $\mathbb{N}$, $\mathbb{R}$, X, and whatever else we might need to refer to. We also assume that the extension is κ-saturated for $\kappa > \max(2^{\aleph_0}, |X|)$ (although we can get away with an often smaller level of saturation; we will discuss this later). Observe that the metric extends to a function $d : X^* \times X^* \to \mathbb{R}^*$ satisfying the axioms in Definition 9.1 above. Thus, (X^*, d) is almost a metric space; it's only defect is that the metric takes values in $\mathbb{R}^*$ rather than $\mathbb{R}$.

We begin by giving a nonstandard characterization of the topological notion of a set being open. First, the standard definition:

Definition 9.3. $\mathcal{O} \subseteq X$ is *open* if for any $a \in \mathcal{O}$, there is $r \in \mathbb{R}^{>0}$ such that $B(a;r) \subseteq \mathcal{O}$.

To give the nonstandard equivalent of open, we first make a very important nonstandard definition:

Definition 9.4. For $a \in X$, the *monad of a* is the set
$$\mu(a) := \{x \in X^* \mid d(a,x) \approx 0\}.$$

Notice that if $X = \mathbb{R}$ is given the euclidean metric, then $\mu(0) = \mu$ is the set of infinitesimals.

Proposition 9.5. *$\mathcal{O} \subseteq X$ is open if and only if, for all $a \in \mathcal{O}$, $\mu(a) \subseteq \mathcal{O}^*$.*

Proof. First assume that $\mathcal{O}$ is open. Fix $a \in \mathcal{O}$. By assumption, there is $r \in \mathbb{R}^{>0}$ such that, for all $x \in X$, if $d(a,x) < r$, then $x \in \mathcal{O}$. Thus, by transfer, if $x \in X^*$ and $d(a,x) < r$, then $x \in \mathcal{O}^*$. In particular, if $x \in \mu(a)$, then $d(a,x) \approx 0$, so $d(a,x) < r$ and hence $x \in \mathcal{O}^*$. Thus, $\mu(a) \subseteq \mathcal{O}^*$.

For the converse, fix $a \in X$. By assumption, $\mu(a) \subseteq \mathcal{O}^*$. Fix $\delta \in \mu^{>0}$. Then if $x \in X^*$ is such that $d(a,x) < \delta$, then $x \in \mu(a)$, whence $x \in \mathcal{O}^*$. Thus, $(\exists \delta \in (\mathbb{R}^{>0})^*)(\forall x \in X^*)(d(a,x) < \delta \to x \in \mathcal{O}^*)$. Now apply the transfer principle to obtain $\delta \in \mathbb{R}^{>0}$ such that, for all $x \in X$, if $d(a,x) < \delta$, then $x \in \mathcal{O}$, whence $B(a;\delta) \subseteq \mathcal{O}$. Since $a \in \mathcal{O}$ was arbitrary, this shows that $\mathcal{O}$ is open. $\square$

Before we state the next corollary, we first note that, by the transfer principle, $B(a;r)^* := \{x \in X^* \mid d(a,x) < r\}$.

Corollary 9.6. *For any $a \in X$ and $r \in \mathbb{R}^{>0}$, $B(a;r)$ is open. (So the terminology* open ball *is appropriate.)*

Proof. Fix $b \in B(a;r)$; we need $\mu(b) \subseteq B(a;r)^*$. Suppose $c \in \mu(b)$, so $d(b,c) \approx 0$. Then, by the (transfer of the) triangle inequality, $d(a,c) \leq d(a,b) + d(b,c) < r$, whence $c \in B(a;r)^*$. $\square$

Exercise 9.7. Prove that $\mu(a)$ is internal if and only if a is an *isolated point* of X, that is, $\{a\}$ is an open set.

Some notation: for $x, y \in X^*$, we write $x \approx y$ to indicate $d(x,y) \approx 0$. So, for $a \in X$, $\mu(a) = \{b \in X^* \mid a \approx b\}$. Also, if $x, y \in X$, then $x \approx y$ if and only if $x = y$.

Exercise 9.8. The relation $\approx$ is an equivalence relation on X^*.

Definition 9.9. $C \subseteq X$ is *closed* if $X \setminus C$ is open.

Corollary 9.10. *C is closed if and only if, whenever $p \in X$ and $q \in C^*$ are such that $p \approx q$, then $p \in C$.*

Proof. $X \setminus C$ is open if and only if for all $p \in X \setminus C$, for all $q \in X^*$ with $p \approx q$, we have $q \in (X \setminus C)^* = X^* \setminus C^*$. $\square$

Digression: Before we continue studying metric space topology, let us just mention briefly how the above set-up generalizes to an arbitrary topological space. First, what is a topological space? A *topological space* is a nonempty set X, together with a collection τ of subsets of X, called the *open subsets of X*, satisfying the following criteria:

- $\emptyset, X \in \tau$;
- If $(U_i \mid i \in I)$ is a family of subsets of X, each of which is in τ, then $\bigcup_{i \in I} U_i \in \tau$;

- If $U, V \in \tau$, then $U \cap V \in \tau$.

So, for example, a metric space, equipped with its collection of open sets (as defined above), is a topological space. There are a plethora of topological spaces not arising from metric spaces and the notion of topological space is central to most areas of mathematics (even logic!). Given a topological space X and $a \in X$, we define the *monad of* a to be $\mu(a) = \bigcap\{U^* \mid U \in \tau,\ a \in U\}$. (Double-check that this agrees with the definition in the metric space context.) With a little more effort, the results we have established in this section (that do not refer to metric notions) hold in the more general context of topological spaces. In fact, this is true of the majority of the results to come (at least the ones that do not mention metric notions).

We return to metric space topology. The following definition, while awkward at first site, is fundamental in topology:

Definition 9.11. $K \subseteq X$ is *compact* if and only if: whenever $(\mathcal{O}_i \mid i \in I)$ is a family of open subsets of X such that $K \subseteq \bigcup_{i \in I} \mathcal{O}_i$, then there are $i_1, \ldots, i_n \in I$ such that $K \subseteq \mathcal{O}_{i_1} \cup \cdots \cup \mathcal{O}_{i_n}$.

In English: every open covering of K has a finite subcover. Observe that, in the above definition of compactness, we can restrict attention to open coverings of K whose index set I has cardinality at most $|K|$. (Why?) Consider the following elegant nonstandard characterization of compactness:

Proposition 9.12 (Robinson's characterization of compactness)**.** *K is compact if and only if, for every $p \in K^*$, there is $q \in K$ with $p \approx q$.*

Proof. First assume that K is compact and yet, towards a contradiction, that there is $p \in K^*$ such that $p \not\approx q$ for every $q \in K$. For each $q \in K$, there is then $r_q \in \mathbb{R}^{>0}$ such that $d(p, q) \geq r_q$. Notice that $\{B(q; r_q) \mid q \in K\}$ is an open cover of K. Thus, since K is compact, there are $q_1, \ldots, q_n \in K$ such that $K \subseteq \bigcup_{i=1}^n B(q_i; r_{q_i})$. In logical terms, the following is true:

$$(\forall x \in K)(d(x, q_1) < r_{q_1} \vee d(x, q_2) < r_{q_2} \vee \cdots \vee d(x, q_n) < r_{q_n}).$$

Applying the transfer principle to the displayed statement, we see that $d(p, q_i) < r_{q_i}$ for some $i \in \{1, \ldots, n\}$, which is a contradiction.

For the converse, assume that K is not compact. Thus, there is an open cover $(\mathcal{O}_i \mid i \in I)$ of K with no finite subcover. For $i \in I$, consider the internal set $A_i := K^* \setminus \mathcal{O}_i^*$. By assumption, each A_i is nonempty and the family $(A_i \mid i \in I)$ has the finite intersection property: indeed, if $A_{i_1} \cap \cdots \cap A_{i_n} = \emptyset$, then for all $x \in K^*$, there is $i \in \{1, \ldots, n\}$ such that $x \in \mathcal{O}_i^*$. By transfer, this would imply that $K \subseteq \bigcup_{j=1}^n \mathcal{O}_{i_j}$, contradicting the choice of the open cover $(\mathcal{O}_i)$. Thus, *by saturation*, there is $p \in \bigcap_{i \in I} A_i$. (Our first serious use of saturation!) Then $p \in K^*$ and $p \not\approx q$ for any $q \in K$: indeed, if $p \approx q$ with $q \in K$, then taking $i \in I$ such that $q \in \mathcal{O}_i$, we would have $p \in \mu(q) \subseteq \mathcal{O}_i^*$, contradicting $p \in A_i$. $\square$

Example 9.13. It is easy to see, using the nonstandard characterization of compactness, that $[a,b]$ is compact. However, if $\epsilon \in \mu^{>0}$, then $a+\epsilon \in (a,b]^*$, but not in the monad of any element of $(a,b]$, whence $(a,b]$ is not compact. Similarly, if $N > \mathbb{N}$, then $N \in [0,\infty)^*$ but not in the monad of any element of $[0,\infty)$, whence $[0,\infty)$ is not compact. Compactness is meant to generalize the notion of closed, bounded intervals $[a,b]$, but we'll soon see that this analogy breaks down in "infinite-dimensional" examples.

Corollary 9.14.

(1) *If $C \subseteq K$, where C is closed and K is compact, then C is compact.*
(2) *If K is compact, then K is closed.*

Proof. For (1), suppose $p \in C^*$; we need to find $q \in C$ such that $p \approx q$. Since $p \in K^*$ as well, by compactness of K, $p \approx q$ for some $q \in K$. By the nonstandard characterization of closed, we have $q \in C$. For (2), suppose that $p \in K^*$, $q \in X$ are such that $p \approx q$; we need $q \in K$. Since K is compact, there is $q' \in K$ such that $p \approx q'$. By Exercise 9.8, $q \approx q'$; since $q, q' \in X$, we have that $q = q'$, whence $q \in K$. □

Here's a question: What is the analog of $\mathbb{R}_{\text{fin}}$ for our metric space X? If we use just the definition of $\mathbb{R}_{\text{fin}}$, then we should make the following definition:

Definition 9.15. The set of *finite points of X^** is

$$X_{\text{fin}} = \{a \in X^* \mid d(a,b) \in \mathbb{R}_{\text{fin}} \text{ for some } b \in X\}.$$

However, by Theorem 1.9, every element of $\mathbb{R}_{\text{fin}}$ is infinitely close to a (standard) real number. This motivates:

Definition 9.16. The set of *nearstandard elements of X^** is

$$X_{\text{ns}} := \{a \in X^* \mid a \approx b \text{ for some } b \in X\}.$$

In other words, $X_{\text{ns}} = \bigcup_{b \in X} \mu(b)$. Some remarks are in order:

Remarks 9.17.

(1) The "for some" in Definition 9.15 can be replaced with "for all," that is,

$$X_{\text{fin}} := \{a \in X^* \mid d(a,b) \in \mathbb{R}_{\text{fin}} \text{ for all } b \in X\}.$$

Indeed, suppose that $a \in X^*$ and $b \in X$ are such that $d(a,b) \in \mathbb{R}_{\text{fin}}$. For any other $c \in X$, we have

$$d(a,c) \leq d(a,b) + d(b,c) \in \mathbb{R}_{\text{fin}} + \mathbb{R} \subseteq \mathbb{R}_{\text{fin}}.$$

(2) It is immediate to see that $X_{\text{ns}} \subseteq X_{\text{fin}}$. Sometimes we have equality; for example, Theorem 1.9 says that $\mathbb{R}_{\text{ns}} = \mathbb{R}_{\text{fin}}$. However, we often have a strict inclusion $X_{\text{ns}} \subsetneq X_{\text{fin}}$. For example, let $X = C([0,1],\mathbb{R})$ from Example 9.2. By transfer, an element of X^* is a function

$f : [0,1]^* \to \mathbb{R}^*$ that satisfies the ϵ-δ definition of continuity for $\epsilon, \delta \in \mathbb{R}^*$. So consider $f : [0,1]^* \to \mathbb{R}^*$ given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq \frac{1}{2} \\ \frac{1}{\epsilon}(x - \frac{1}{2}) & \text{if } \frac{1}{2} \leq x \leq \frac{1}{2} + \epsilon \\ 1 & \text{otherwise,} \end{cases}$$

where $\epsilon \in \mu^{>0}$. Then $f \in X^*$, but since f makes an "appreciable" jump in an infinitesimal time period, f cannot be infinitely close to a standard, continuous function $g : [0,1] \to \mathbb{R}$. We will later clarify exactly when $X_{\mathrm{ns}} = X_{\mathrm{fin}}$.

(3) Suppose $p \in X_{\mathrm{ns}}$. Then there is a *unique* $q \in X$ such that $p \approx q$: indeed, if $p \approx q$ and $p \approx q'$, where $q, q' \in X$, then $q \approx q'$, so $q = q'$. In analogy with earlier in these notes, we call this unique q the *standard part of* p, denoted $\mathrm{st}(p)$.

Robinson's characterization of compactness can now be phrased as:

Corollary 9.18. *X is compact if and only if $X^* = X_{\mathrm{ns}}$.*

Definition 9.19. $B \subseteq X$ is *bounded* if there is $a \in X$ and $r \in \mathbb{R}^{>0}$ such that $B \subseteq B(a; r)$.

Proposition 9.20. *$B \subseteq X$ is bounded if and only if $B^* \subseteq X_{\mathrm{fin}}$.*

Proof. Suppose that B is bounded, say $B \subseteq B(a; r)$. Then $B^* \subseteq B(a; r)^*$; clearly $B(a; r)^*$ is contained in X_{fin}. Conversely, suppose that $B^* \subseteq X_{\mathrm{fin}}$. Fix $a \in X$ and $N > \mathbb{N}$. Then the following is true in the nonstandard universe:

$$(\exists N \in \mathbb{N}^*)(\forall x \in B^*)(d(a, x) < N).$$

Now apply transfer. □

In particular, X is bounded if and only if $X^* = X_{\mathrm{fin}}$.

Corollary 9.21. *If $K \subseteq X$ is compact, then K is bounded.*

Proof. This follows from the inclusions $K^* \subseteq X_{\mathrm{ns}} \subseteq X_{\mathrm{fin}}$. □

Definition 9.22. X is a *Heine-Borel* (or *proper*) metric space if, for all $K \subseteq X$, we have K is compact if and only if K is closed and bounded.

The name Heine-Borel metric space comes from the *Heine-Borel Theorem* (to be proven below), which states that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded. We have already seen that compact sets are always closed and bounded, so the meat of the definition is the other implication.

Proposition 9.23. *X is a* Heine-Borel *metric space if and only if $X_{\mathrm{ns}} = X_{\mathrm{fin}}$.*

Proof. First suppose that X is a Heine-Borel metric space and suppose that $a \in X_{\text{fin}}$; we need $a \in X_{\text{ns}}$. Fix $b \in X$; then $d(a,b) \in \mathbb{R}_{\text{fin}}$, say $d(a,b) < r$ with $r \in \mathbb{R}^{>0}$. Then $a \in \bar{B}(b;r)^*$. Since X is Heine-Borel, we have $\bar{B}(b;r)$ is compact, whence $a \approx c$ for some $c \in \bar{B}(b;r)$. In particular, $a \in X_{\text{ns}}$.

Conversely, suppose that $X_{\text{ns}} = X_{\text{fin}}$. Suppose that $K \subseteq X$ is closed and bounded; we need K to be compact. Fix $a \in K^*$; we need $a \approx b$ for some $b \in K$. Since K is bounded, we have $K^* \subseteq X_{\text{fin}}$, whence $a \in X_{\text{fin}} = X_{\text{ns}}$. Thus, there is (unique) $b \in X$ such that $a \approx b$. It remains to verify that $b \in K$; but this follows immediately from the fact that K is closed (and the nonstandard characterization of closed). □

Corollary 9.24. $\mathbb{R}^n$ *is a Heine-Borel metric space.* $C([0,1],\mathbb{R})$ *is not a Heine-Borel metric space.*

We can define continuity between metric spaces. Suppose that Y is also a metric space and $f : X \to Y$ is a function. For $p \in X$, we say that f is *continuous at* p if whenever $\mathcal{O} \subseteq Y$ is open and $f(p) \in \mathcal{O}$, then there is an open $\mathcal{O}' \subseteq X$ such that $f(\mathcal{O}') \subseteq f(\mathcal{O})$. We say that f is *continuous* if f is continuous at p for all $p \in X$.

The astute observer will notice that this is not the direct translation of continuity for functions on $\mathbb{R}$. However, the following exercise will make them feel better:

Exercise 9.25. The following are equivalent:

(1) f is continuous at p;
(2) For all $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $q \in X$, if $d(p,q) < \delta$, then $d(f(p), f(q)) < \epsilon$;
(3) $f(\mu(p)) \subseteq \mu(f(p))$, that is, if $q \approx p$, then $f(q) \approx f(p)$.

We use the above definition for continuity as it makes sense in an arbitrary topological space and not just for metric spaces. The equivalence of (1) and (3) in the previous exercise will still hold in this more general context.

Proposition 9.26. *Suppose that* $f : X \to Y$ *is continuous and* $K \subseteq X$ *is compact. Then* $f(K) \subseteq Y$ *is compact.*

Proof. Suppose $y \in f(K)^*$; we need $y \in f(K)_{\text{ns}}$. By transfer, we have $y = f(x)$ for some $x \in K^*$. Since K is compact, $\text{st}(x)$ exists and belongs to K. Since f is continuous at $\text{st}(x)$, $y = f(x) \in \mu(f(\text{st}(x)))$, so $y \in f(K)_{\text{ns}}$. □

Exercise 9.27. Suppose that $f : X \to Y$ is a function.

(1) Define what it means for f to be uniformly continuous. Then state and prove a nonstandard characterization of uniform continuity.
(2) Suppose that f is continuous and X is compact. Prove that f is uniformly continuous.

For the purpose of the next exercise, define $X_{\text{inf}} := X^* \setminus X_{\text{fin}}$. A (not necessarily continuous) function $f : X \to Y$ is said to be *proper* if $f^{-1}(K) \subseteq X$ is compact for every compact $K \subseteq Y$.

Exercise 9.28. Suppose that X and Y are Heine Borel metric spaces and $f : X \to Y$ is continuous. Prove that f is proper if and only if $f(X_{\text{inf}}) \subseteq Y_{\text{inf}}$.

We can also bring the notions of *sequences* and *convergence of sequences* into the metric space setting. For example, a sequence (a_n) from X *converges to* $a \in X$ if and only if, for every $\epsilon \in \mathbb{R}^{>0}$, there is $m \in \mathbb{N}$ such that, for all $n \in \mathbb{N}$, if $n \geq m$, then $d(a_n, a) < \epsilon$.

Here is the metric space version of Bolzano-Weierstrauss:

Theorem 9.29. *If X is a compact metric space and (a_n) is a sequence in X, then a_n has a convergent subsequence.*

Proof. Fix $N > \mathbb{N}$. Then $a_N \in X^* = X_{\text{ns}}$. Then $\text{st}(a_N)$ is a limit point of (a_n). □

Definition 9.30. X is a *complete metric space* if every Cauchy sequence in X converges.

Corollary 9.31. *Compact metric spaces are complete.*

Proof. Suppose that (a_n) is a Cauchy sequence in X, so $a_M \approx a_N$ for all $M, N > \mathbb{N}$. Since X is compact, $a_N \in X_{\text{ns}}$ for all $N > \mathbb{N}$. Thus, if $L = \text{st}(a_N)$ for $N > \mathbb{N}$, then $a_M \approx L$ for all $M > \mathbb{N}$, whence (a_n) converges to L. □

Exercise 9.32. Suppose that X is complete and $C \subseteq X$ is closed. Prove that C is also complete.

In order to explain the nonstandard characterization of completeness, it is convenient at this point to introduce another important set of points in X^*:

Definition 9.33. The set of *pre-nearstandard points of X^** is

$$X_{\text{pns}} := \{a \in X^* \mid \text{ for each } \epsilon \in \mathbb{R}^{>0}, \text{ there is } b \in X \text{ such that } d(a, b) < \epsilon\}.$$

Immediately, we see that $X_{\text{ns}} \subseteq X_{\text{pns}} \subseteq X_{\text{fin}}$.

Theorem 9.34. *X is complete if and only if $X_{\text{ns}} = X_{\text{pns}}$.*

Proof. First suppose that X is complete and $p \in X_{\text{pns}}$. Then, for every $n \geq 1$, there is $q_n \in X$ such that $d(p, q_n) < \frac{1}{n}$. It follows that (q_n) is Cauchy, whence converges to $q \in X$. It follows that $p \approx q$, whence $p \in X_{\text{ns}}$.

Towards the converse, suppose that $X_{\text{ns}} = X_{\text{pns}}$ and suppose that (x_n) is Cauchy. Fix $N > \mathbb{N}$; it suffices to show that $x_N \in X_{\text{ns}}$. If not, then $x_N \notin X_{\text{pns}}$, whence there is $\epsilon \in \mathbb{R}^{>0}$ such that $d(x_N, q) \geq \epsilon$ for all $q \in X$. In particular, $d(x_N, x_n) \geq \epsilon$ for all $n \in \mathbb{N}$. But (x_n) is Cauchy, so for some $n \in \mathbb{N}$ big enough, $d(x_N, x_n) < \epsilon$, a contradiction. □

Corollary 9.35. *If X is Heine-Borel, then X is complete.*

The following theorem on "remoteness" will prove useful later in these notes:

Theorem 9.36. *Suppose that $(p_n \mid n \in \mathbb{N}^*)$ is an internal sequence from X^*. Suppose that $r \in \mathbb{R}^{>0}$ is such that $d(p_m, p_n) \geq r$ for all distinct $m, n \in \mathbb{N}$. Then $p_n \notin X_{\mathrm{ns}}$ for some $n \in \mathbb{N}^*$.*

Proof. Suppose, towards a contradiction, that $p_n \in X_{\mathrm{ns}}$ for all $n \in \mathbb{N}^*$. For $n \in \mathbb{N}$ set $q_n := \mathrm{st}(p_n)$. We then get the nonstandard extension of (q_n), namely $(q_n \mid n \in \mathbb{N}^*)$. We must be careful here: just because $p_n \approx q_n$ for all $n \in \mathbb{N}$ does not imply that $p_n \approx q_n$ for all $n \in \mathbb{N}^*$ (as the relation $\approx$ is external). Nevertheless, the sequence $(d(p_n, q_n) \mid n \in \mathbb{N}^*)$ is internal and infinitesimal for $n \in \mathbb{N}$. Thus, by the Infinitesimal Prolongation Theorem, there is $N > \mathbb{N}$ such that $d(p_n, q_n) \approx 0$ for all $n \leq N$. Fix $M > \mathbb{N}$ with $M < N$ and set $q := \mathrm{st}(p_M) = \mathrm{st}(q_M)$ (which is possible by our standing assumption). Thus, there is a subsequence (q_{n_k}) converging to q. Choose $n_0 \in \mathbb{N}$ such that, for $j, k \geq n_0$: $d(q_{n_j}, q_{n_k}) < \frac{r}{2}$. It follows that $d(p_{n_j}, p_{n_k}) < r$, a contradiction. $\square$

9.2. **More about continuity.**

Definition 9.37. Suppose that $f : X^* \to Y^*$ is a function.

(1) f is *S-continuous* if, for all $x, x' \in X^*$, if $x \approx x'$, then $f(x) \approx f(x')$.
(2) f is *$\epsilon\delta$-continuous* if, for all $p \in X^*$ and $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $q \in X^*$, if $d(p, q) < \delta$, then $d(f(p), f(q)) < \epsilon$.

The important point in (2) is that both ϵ and δ are standard. It is easy to see that $\epsilon\delta$-continuity implies S-continuity. The converse need not hold:

Exercise 9.38. Consider $f : \mathbb{R}^* \to \mathbb{R}^*$ defined by $f(x) = 0$ if $x \approx 0$ while $f(x) = 1$ otherwise. Show that f is S-continuous but not $\epsilon\delta$-continuous.

Observe that the function f in the previous exercise is external. For internal functions, the above notions coincide:

Proposition 9.39. *Suppose that $f : X^* \to Y^*$ is internal and S-continuous. Then f is $\epsilon\delta$-continuous.*

Proof. Suppose, towards a contradiction, that there is $\epsilon \in \mathbb{R}^{>0}$ and $p \in X^*$ such that, for every $\delta \in \mathbb{R}^{>0}$, there is $q \in X^*$ such that $d(p, q) < \delta$ while $d(f(p), f(q)) \geq \epsilon$. By saturation (how?), there is $q \in X^*$ such that $d(p, q) \approx 0$ while $d(f(p), f(q)) \geq \epsilon$, contradicting S-continuity. $\square$

The following construction is crucial in defining standard continuous functions from internal functions.

Theorem 9.40. *Suppose X and Y are metric spaces with X compact. Suppose that $f : X^* \to Y^*$ is an internal, S-continuous function. Suppose further that $f(p) \in Y_{\mathrm{ns}}$ for each $p \in X$. Define $F : X \to Y$ by $F(p) = \mathrm{st}(f(p))$. Then F is continuous and $F(p) \approx f(p)$ for all $p \in X^*$.*

Proof. Fix $p \in X$; we show that F is continuous at p. Fix $\epsilon \in \mathbb{R}^{>0}$. By Proposition 9.39, there is $\delta \in \mathbb{R}^{>0}$ witnessing that f is $\epsilon\delta$-continuous for

$\frac{\epsilon}{2}$. Suppose $q \in X$ is such that $d(p,q) < \delta$. Then $d(f(p), f(q)) < \frac{\epsilon}{2}$. Since $f(p) \approx F(p)$ and $f(q) \approx F(q)$, this shows that $d(F(p), F(q)) < \epsilon$.

Now suppose that $p \in X^*$; we need $F(p) \approx f(p)$. Let $p' := \mathrm{st}(p)$; this is possible since X is compact. Then $f(p) \approx f(p')$ by S-continuity of f. Meanwhile, $F(p') \approx f(p')$ by definition of F and $F(p') \approx F(p)$ by continuity of F. □

We recall the definition of equicontuity, this time in the metric space setting: a sequence of functions (f_n) from X to Y is *equicontinuous* if, for all $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that, for all $p, q \in X$ and $n \in \mathbb{N}$: if $d(p,q) < \delta$, then $d(f_n(p), f_n(q)) < \epsilon$.

Exercise 9.41. Suppose that X and Y are metric spaces and X is compact. Set $Z := C(X,Y)$ to be the space of continuous functions from X to Y. Define a function d on Z by setting $d(f,g) := \sup_{x \in X} d(f(x), g(x))$.

(1) Show that d is a metric on Z. (In particular, this means showing that the supremum is never infinite.)
(2) Show that (f_n) converges to f (in the sense of the metric space Z) if and only if (f_n) converges to f uniformly.

If X is a compact metric space and $f_n : X \to \mathbb{R}$ is a continuous function for each $n \in \mathbb{N}$, we say that the sequence (f_n) is *uniformly bounded* if there is $M \in \mathbb{R}$ such that $|f_n(x)| \leq M$ for all $x \in X$ and all $n \in \mathbb{N}$.

Corollary 9.42 (Arzela-Ascoli). *If X is a compact metric space and (f_n) is a uniformly bounded, equicontinuous sequence of functions from X to $\mathbb{R}$, then (f_n) has a uniformly convergent subsequence.*

Proof. Fix $N > \mathbb{N}$. As in an earlier part of the notes, $f_N : X^* \to \mathbb{R}^*$ is S-continuous (and internal). Since (f_n) is uniformly bounded, it follows that $f_N(X) \subseteq \mathbb{R}_{\mathrm{ns}}$.Define $F : X \to \mathbb{R}$ (as in the last theorem) by setting $F(p) := \mathrm{st}(f_N(p))$. By the last exercise, it suffices to show that F is a limit point of (f_n); to do this, we will show that $F \approx f_N$ as elements of $C(X,\mathbb{R})^*$. Well, by transfer, $d(F, f_N) < \epsilon$ if and only if $d(F(p), f_N(p)) < \epsilon$ for all $p \in X^*$; this follows immediately from the conclusion of the previous theorem. □

9.3. **Compact maps.** We now discuss an important class of functions that will appear later in the functional analysis section. Once again, X and Y are metric spaces.

Definition 9.43. $f : X \to Y$ is *compact* if, for every bounded $B \subseteq X$, we have a compact $K \subseteq Y$ such that $f(B) \subseteq K$.

In other words, f is compact if and only if $\overline{f(B)}$ is compact.

Theorem 9.44. *$f : X \to Y$ is compact if and only if $f(X_{\mathrm{fin}}) \subseteq Y_{\mathrm{ns}}$.*

Proof. First suppose that f is compact. Fix $p \in X_{\text{fin}}$; we need $f(p) \in Y_{\text{ns}}$. Well, $p \in B := B(a;r)$ for some $a \in X$ and $r \in \mathbb{R}^{>0}$, whence $f(p) \in \overline{f(B)}^* \subseteq Y_{\text{ns}}$ since $\overline{f(B)}$ is compact.

Conversely, suppose that $f(X_{\text{fin}}) \subseteq Y_{\text{ns}}$. Fix $B \subseteq X$ bounded; we must show that $\overline{f(B)}$ is compact. Take $q \in \overline{f(B)}^*$; we must find $q' \in \overline{f(B)}$ such that $q \approx q'$. Fix $\epsilon \in \mu^{>0}$; by transfer, there is $y \in f(B)^*$ such that $d(q,y) < \epsilon$. Write $y = f(x)$ for $x \in B^*$. By assumption, $f(x) \in Y_{\text{ns}}$, so $f(x) \approx q'$ for some $q' \in Y$. It remains to show that $q' \in \overline{f(B)}$. Fix $\delta \in \mathbb{R}^{>0}$. By assumption, there is $z \in f(B)^*$ such that $d(q,z) < \delta$, whence it follows that $d(q',z) < \delta$. Applying transfer to this last fact, we see that there is $z \in f(B)$ such that $d(q',z) < \delta$. $\square$

Corollary 9.45. *Suppose that $f : X \to Y$ is a function. If Y is compact, then f is compact.*

Corollary 9.46. *Suppose that (f_n) is a sequence of compact functions from X to Y. Further assume that Y is complete and that (f_n) converges uniformly to f. Then f is compact.*

Proof. Suppose $x \in X_{\text{fin}}$; we need $f(x) \in Y_{\text{ns}}$. Since Y is complete, it suffices to prove that $f(x) \in Y_{\text{pns}}$. Fix $\epsilon \in \mathbb{R}^{>0}$. Fix $m \in \mathbb{N}$ such that $d(f_m(p), f(p)) < \frac{\epsilon}{2}$ for all $p \in X$. By transfer, $d(f_m(x), f(x)) < \frac{\epsilon}{2}$. Since f_m is compact, we have $f_m(x) \in Y_{\text{ns}}$, say $f_m(x) \approx y$ with $y \in Y$. It follows that $d(f(x), y) < \epsilon$. Since ϵ was arbitrary, this shows that $f(x) \in Y_{\text{pns}}$. $\square$

9.4. **Problems.** You may assume any level of saturation that you need in any given problem.

Problem 9.1. Suppose that X is a metric space and A is a subset of X. The *interior* of A, denoted A°, is defined by

$$A^\circ := \{x \in A \mid \text{there exists } r \in \mathbb{R}^{>0} \text{ such that } B(x,r) \subseteq A\}.$$

(1) Show that A is open iff $A = A^\circ$. (Standard reasoning)
(2) Show that $A^\circ = \bigcup\{\mathcal{O} \mid \mathcal{O} \text{ is open and } \mathcal{O} \subseteq A\}$. (Standard reasoning)
(3) Show that, for any $x \in X$, we have $x \in A^\circ$ iff $y \in A^*$ for any $y \in X^*$ with $y \approx x$.

Problem 9.2. Suppose that X is a metric space and A is a subset of X. The *closure* of A, denoted $\overline{A}$. is defined by

$$\overline{A} := \{x \in X \mid \text{for any } r \in \mathbb{R}^{>0}, \text{ there is } a \in A \text{ such that } d(x,a) < r\}.$$

(1) Show that $\overline{A} = \{x \in X \mid \text{there is } (a_n) \text{ from } A \text{ such that } a_n \to x\}$. (Standard reasoning)
(2) Show that $\overline{A} = \bigcap\{F \mid F \text{ is closed and } A \subseteq F\}$. (Standard reasoning)
(3) Show that A is closed iff $A = \overline{A}$. (Standard reasoning)
(4) Show that, for any $x \in X$, we have $x \in \overline{A}$ iff there is $y \in A^*$ such that $x \approx y$.

Problem 9.3. Let $C \subseteq \mathbb{R}_{\text{fin}}$ be internal. Define $\text{st}(C) := \{st(c) \mid c \in C\} \subseteq \mathbb{R}$. Prove that $\text{st}(C)$ is closed.

Problem 9.4.

(1) Suppose $f : X \to Y$ is continuous. Show that $f : X^* \to Y^*$ is $*$-continuous.
(2) Suppose that $f : X^* \to Y^*$ is $\epsilon\delta$-continuous. Show that f is S-continuous.
(3) Consider the function $f : \mathbb{R}^* \to \mathbb{R}^*$ defined by
$$f(x) = \begin{cases} 0 & \text{if } x \approx 0 \\ 1 & \text{otherwise.} \end{cases}$$
Show that f is S-continuous, but not $\epsilon\delta$-continuous.
(4) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Show that $f : \mathbb{R}^* \to \mathbb{R}^*$ is $*$-continuous, but not S-continuous.
(5) Fix $\alpha \in \mu^{>0}$. Consider the function $f : \mathbb{R}^* \to \mathbb{R}^*$ defined by
$$f(x) = \begin{cases} \alpha \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$
Show that f is S-continuous at 0 but not $*$-continuous at 0.

Problem 9.5. Suppose that $A \subseteq X^*$ is internal. Let
$$\text{st}(A) = \{x \in X \mid x \approx y \text{ for some } y \in A\}.$$

(1) Show that $\text{st}(A)$ is closed.
(2) Suppose that $A \subseteq X_{\text{ns}}$ is internal. Show that $\text{st}(A)$ is compact.
(3) Suppose that $A \subseteq X$. Show that $\overline{A} = \text{st}(A^*)$.
(4) We say that $A \subseteq X$ is *relatively compact* if $\overline{A}$ is compact. Show that $A \subseteq X$ is relatively compact iff $A^* \subseteq X_{\text{ns}}$.

Problem 9.6. . Let $C([0,1], \mathbb{R})$ denote the set of all continuous functions from $[0,1]$ to $\mathbb{R}$. For $f, g \in C([0,1], \mathbb{R})$, set
$$d(f,g) := \sup\{|f(x) - g(x)| \mid x \in [0,1]\}.$$

(1) Show that $C([0,1], \mathbb{R})^*$ is the set of $*$-continuous functions from $[0,1]^*$ to $\mathbb{R}^*$.
(2) Suppose $f \in C([0,1], \mathbb{R})^*$. Show that $f \in C([0,1], \mathbb{R})_{\text{ns}}$ iff f is S-continuous and $f(x) \in \mathbb{R}_{\text{fin}}$ for all $x \in [0,1]$.
(3) Show that $C([0,1], \mathbb{R})_{\text{ns}} \subsetneq C([0,1], \mathbb{R})_{\text{fin}}$.
(4) Convince yourself that the results of this problem remains true when $[0,1]$ is replaced by any compact metric space X.

Problem 9.7. (Arzela-Ascoli Theorem-reformulated) Suppose that $\mathcal{F} \subseteq C([0,1], \mathbb{R})$. Show that the following are equivalent:

(1) $\mathcal{F}$ is relatively compact;
(2) $\mathcal{F}^* \subseteq C([0,1], \mathbb{R})_{\text{ns}}$;

(3) $\mathcal{F}$ is equicontinuous and, for all $x \in [0,1]$, the set
$$\mathcal{F}_x := \{f(x) \mid f \in \mathcal{F}\}$$
is relatively compact.

(The notion of equicontinuity is exactly as in Section 3, namely, $\mathcal{F}$ is equicontinuous iff for every $\epsilon \in \mathbb{R}^{>0}$, there is $\delta \in \mathbb{R}^{>0}$ such that for all $x, y \in [0,1]$ and all $f \in \mathcal{F}$, if $|x-y| < \delta$, then $|f(x) - f(y)| < \epsilon$. You will need to show again that $\mathcal{F}$ is equicontinuous iff each $f \in \mathcal{F}^*$ is S-continuous.)

By (4) of Problem 9.6, the equivalence in the previous problem remains true when $[0,1]$ is replaced by any compact metric space.

10. Banach Spaces

In this section, $\mathbb{F}$ denotes one of the two fields $\mathbb{R}$ or $\mathbb{C}$. Let's say a word about $\mathbb{C}^*$. By transfer, the elements of $\mathbb{C}^*$ are of the form $z = x + iy$ for $x, y \in \mathbb{R}^*$ and then $|z| = \sqrt{x^2 + y^2}$. It is then straightforward to verify that $\mathbb{C}_{\text{fin}} = \{z \in \mathbb{C}^* \mid |z| \in \mathbb{R}_{\text{fin}}\} = \{x + iy \mid x, y \in \mathbb{R}_{\text{fin}}\} = \mathbb{C}_{\text{ns}}$ since $\mathbb{R}_{\text{fin}} = \mathbb{R}_{\text{ns}}$. If $z = x + iy \in \mathbb{C}_{\text{ns}}$, then $\operatorname{st}(z) = \operatorname{st}(x) + i \operatorname{st}(y)$.

10.1. **Normed spaces.**

Definition 10.1. If V is a vector space over $\mathbb{F}$, then a *norm* on V is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying, for all $x, y \in V$ and $\alpha \in \mathbb{F}$:

(1) $\|x\| \geq 0$;
(2) $\|x\| = 0$ if and only if $x = 0$;
(3) $\|\alpha x\| = |\alpha| \cdot \|x\|$;
(4) (Triangle Inequality) $\|x + y\| \leq \|x\| + \|y\|$.

A *normed space* is a vector space equipped with a norm.

Example 10.2.

(1) For any $p \geq 1$, the *p-norm* on $\mathbb{F}^n$ is given by $\|x\|_p := \sqrt[p]{\sum_{i=1}^n |x_i|^p}$.
(2) There is an infinitary analog of the previous example. Fix $p \geq 1$ and set ℓ_p to be the set of all infinite sequences (α_n) from $\mathbb{F}$ such that $\sum_{n=1}^\infty |\alpha_n|^p < \infty$. Then ℓ_p is a vector space over $\mathbb{F}$ and we define the p-norm on ℓ_p in the exact same way: $\|(\alpha_n)\|_p := \sqrt[p]{\sum_{n=1}^\infty |\alpha_n|^p}$.
(3) Fix a compact metric space X and set $C(X, \mathbb{F}) := \{f : X \to \mathbb{F} \mid f \text{ is continuous}\}$. Then $C(X, \mathbb{F})$ is a (usually infinite-dimensional) vector space over $\mathbb{F}$ and $\|f\| := \sup_{x \in X} |f(x)|$ defines a norm on $C(X, \mathbb{F})$.

Exercise 10.3. Suppose that $(V, \|\cdot\|)$ is a normed space. Define $d : V \times V \to \mathbb{R}$ by $d(x, y) := \|x - y\|$. Then d is a metric on V and $d(x, y) = d(x - y, 0)$ for any $x, y \in V$.

We will always treat a normed space as a metric space as in the previous exercise. A normed space is called a *Banach space* if the associated metric is complete.

Exercise 10.4. Verify that all of the normed spaces from Example 10.2 are Banach spaces. (Hints: Don't forget our nonstandard characterization of completeness. Also, for showing that $C(X, \mathbb{F})$ is a Banach space, don't forget about our nonstandard characterization of $C(X, \mathbb{F})_{\mathrm{ns}}$ from Problem 9.6.)

Until otherwise stated, we fix normed spaces V and W; we write d for both of the associated metrics on V and W. For $x \in V^*$, we say that x is *infinitesimal* if $x \approx 0$, that is, $d(x, 0) \in \mu$ (equivalently, $\|x\| \in \mu$). It follows immediately that $x \approx y$ if and only if $x - y$ is infinitesimal.

Lemma 10.5. *If $x, y \in V^*$ and $x \approx y$, then $\|x\| \approx \|y\|$. (The converse fails miserably!)*

Proof. We may suppose that $\|x\| \leq \|y\|$ Write $y = x + (y - x)$. Then $\|y\| \leq \|x\| + \|y - x\| \approx \|x\|$ since $y - x$ is infinitesimal. Thus, $\|x\| \approx \|y\|$. $\square$

Exercise 10.6.

(1) If $\alpha \in \mathbb{F}_{\mathrm{fin}}$ and $x, y \in V^*$ are such that $x \approx y$, show that $\alpha x \approx \alpha y$.
(2) Prove that the addition and scalar multiplication maps $+ : V \times V \to V$ and $\cdot : \mathbb{F} \times V \to V$ are continuous (with respect to the metric d). Please use the nonstandard characterization of continuity.

10.2. **Bounded linear maps.**

Proposition 10.7. *Suppose that $T : V \to W$ is a linear transformation and T is continuous at some $x_0 \in V$. Then T is uniformly continuous.*

Proof. We use the nonstandard characterization of uniform continuity: suppose $x, y \in V^*$ and $x \approx y$. We show that $Tx \approx Ty$. Well, $x_0 + (x - y) \approx x_0$, so by the continuity of T at x_0, we have $T(x_0 + x - y) \approx T(x_0)$. Thus, $T(x_0) + T(x) - T(y) \approx T(x_0)$, whence $T(x) \approx T(y)$. (We have used the transfer principle to infer that the nonstandard extension of T is also linear.) $\square$

Exercise 10.8. Suppose that $T : V \to W$ is a linear transformation that is continuous. Prove that $\ker(T) := \{x \in V \mid T(x) = 0\}$ is a *closed* subspace of V.

Definition 10.9. We say that a linear transformation $T : V \to W$ is bounded if there is $M \in \mathbb{R}^{>0}$ such that $\|Tx\| \leq M\|x\|$ for all $x \in V$.

The terminology in the above definition corresponds to the next fact:

Proposition 10.10. *$T : V \to W$ is bounded if and only if $\{T(x) \mid \|x\| = 1\}$ is a bounded subset of W.*

Proof. Let $A := \{T(x) \mid \|x\| = 1\}$. For the ($\Rightarrow$) direction, if $\|Tx\| \leq M\|x\|$ for all $x \in V$, then A is contained in the closed ball around 0 (in W) of radius M. Conversely, suppose A is contained in the closed ball around 0 of radius M. We claim that $\|Tx\| \leq M\|x\|$ for all $x \in V$. Indeed, for $x \in V \setminus \{0\}$, $\|\frac{1}{\|x\|}x\| = 1$, so $\|T(\frac{1}{\|x\|}x)\| \leq M$, whence $\|Tx\| \leq M\|x\|$. $\square$

If $T : V \to W$ is bounded, set $\|T\| := \sup\{\|T(x)\| \mid \|x\| = 1\}$, which is a real number by the previous proposition. Taking $M = \|T\|$ in the proof of the ($\Leftarrow$) direction of the previous proposition, we obtain the following:

Corollary 10.11. $\|Tx\| \leq \|T\|\|x\|$ *for all* $x \in V$.

Definition 10.12. We set $\mathcal{B}(V, W)$ to be the set of all bounded linear transformations from V to W.

Exercise 10.13. Show that $\mathcal{B}(V, W)$ is a vector subspace of the set of all linear transformations from V to W. Further show that $\mathcal{B}(V, W)$ is a normed space (with the above definition of $\|T\|$). If $V = W$, prove that whenever $T, U \in \mathcal{B}(V, V)$, then $T \circ U \in \mathcal{B}(V, V)$ and $\|T \circ U\| \leq \|T\| \cdot \|U\|$. (This shows that $\mathcal{B}(V, V)$ is a *normed algebra*.)

Theorem 10.14. *Suppose that* $T : V \to W$ *is linear. The following are equivalent:*

(1) T *is continuous at* x_0 *for some* $x_0 \in V$*;*
(2) T *is uniformly continuous;*
(3) T *is bounded;*
(4) $T(V_{\text{fin}}) \subseteq W_{\text{fin}}$*;*
(5) $T(V_{\text{ns}}) \subseteq W_{\text{ns}}$.

Proof. The equivalence of (1) and (2) follows from Proposition 10.7 and (2) $\Rightarrow$ (5) is clear. Suppose that T is not bounded. Let

$$X = \{n \in \mathbb{N}^* \mid \|T(x)\| > n \text{ for some } x \in V^* \text{ with } \|x\| = 1\},$$

an internal subset of $\mathbb{N}^*$ that contains $\mathbb{N}$ by assumption. Thus, by overflow, there is $N \in X \setminus \mathbb{N}$. Choose $x \in V^*$ with $\|Tx\| > N$ and $\|x\| = 1$. At this point we have established the implication (4) $\Rightarrow$ (3). Set $y := \frac{1}{\|T(x)\|}x$, so $\|y\| < \frac{1}{N}$, whence $y \approx 0$. But $\|T(y)\| = 1$, so $T(y) \not\approx y$, that is, T is not continuous. This proves the direction (2) $\Rightarrow$ (3). Set $z := \frac{1}{\sqrt{\|Tx\|}}x$; since $\|z\| \in \mu$, we have $z \in V_{\text{ns}}$. But $\|Tz\| = \frac{1}{\sqrt{\|Tx\|}}\|Tx\| \in \mathbb{R}_{\text{inf}}$, so $Tz \notin W_{\text{ns}}$; this shows (5) $\Rightarrow$ (3). For (3) $\Rightarrow$ (1), if $x \approx 0$, then $\|Tx\| \leq \|T\| \cdot \|x\| \approx 0$, so $Tx \approx 0$ and T is continuous at 0. For (3) $\Rightarrow$ (4), assume that T is bounded and fix $x \in V_{\text{fin}}$. Then $\|T(x)\| \leq \|T\| \cdot \|x\| \in \mathbb{R}_{\text{fin}}$, whence $T(x) \in W_{\text{fin}}$. $\square$

10.3. **Finite-dimensional spaces and compact linear maps.** We now aim to understand what happens for finite-dimensional normed spaces. First, some lemmas.

Lemma 10.15. *Suppose that* $x_1, \ldots, x_n \in V_{\text{fin}}$*,* $\alpha_1, \ldots, \alpha_n \in \mathbb{F}^*$ *are such that* $\alpha_1, \ldots, \alpha_n \approx 0$*. Then* $\alpha_1 x_1 + \cdots + \alpha_n x_n \approx 0$.

Proof. Immediate from the triangle inequality. $\square$

Lemma 10.16. *Suppose that* $x_1, \ldots, x_n \in V$ *are linearly independent and* $\alpha_1, \ldots, \alpha_n \in \mathbb{F}^*$ *are such that* $\alpha_1 x_1 + \cdots + \alpha_n x_n \in V_{\text{fin}}$*. Then* $\alpha_i \in \mathbb{F}_{\text{fin}}$ *for all* $i = 1, \ldots, n$.

Proof. Without loss of generality (by renumbering if necessary), we may assume that $\max\{|\alpha_1|, \ldots, |\alpha_n|\} = |\alpha_1|$. Suppose, towards a contradiction, that $\alpha_1 \notin \mathbb{F}_{\text{fin}}$. Then $\|x_1 + \frac{\alpha_2}{\alpha_1}x_2 + \cdots + \frac{\alpha_n}{\alpha_1}x_n\| = \frac{1}{|\alpha_1|}\|\alpha_1 x_1 + \cdots + \alpha_n x_n\| \approx 0$ since $|\frac{1}{\alpha_1}| \approx 0$ and $\|\alpha_1 x_1 + \cdots + \alpha_n x_n\| \in \mathbb{R}_{\text{fin}}$. Since $|\alpha_i| \leq |\alpha_1|$ for each $i = 1, \ldots, n$, $\text{st}(|\frac{\alpha_i}{\alpha_1}|)$ exists. By the previous lemma, we have

$$0 \approx x_1 + \frac{\alpha_2}{\alpha_1}x_2 + \cdots + \frac{\alpha_n}{\alpha_1}x_n \approx x_1 + \text{st}(\frac{\alpha_2}{\alpha_1})x_2 + \cdots + \text{st}(\frac{\alpha_n}{\alpha_1})x_n.$$

Since the term on the right of the above display is standard, it equals 0. This contradicts the fact that $x_1, \ldots, x_n$ are linearly independent. □

Corollary 10.17. *If V is finite-dimensional, then $V_{\text{fin}} = V_{\text{ns}}$*

Proof. Let $\{x_1, \ldots, x_n\}$ be a basis for V. By transfer, every element of V^* is a $\mathbb{F}^*$-linear combination of $x_1, \ldots, x_n$. If $x = \alpha_1 x_1 + \ldots + \alpha_n x_n \in V_{\text{fin}}$, then each $\alpha_i \in \mathbb{F}_{\text{fin}} = \mathbb{F}_{\text{ns}}$ by the previous lemma, whence x is infinitely close to $\text{st}(\alpha_1)x_1 + \cdots + \text{st}(\alpha_n)x_n \in V$. □

Corollary 10.18. *If V is finite-dimensional, then V is a Banach space.*

Corollary 10.19. *If $T : V \to W$ is a linear transformation and V is finite-dimensional, then T is bounded.*

Proof. By Theorem 10.14, it suffices to prove that $T(V_{\text{fin}}) \subseteq W_{\text{fin}}$. Fix $x = \alpha_1 x_1 + \cdots + \alpha_n x_n \in V_{\text{fin}}$, so each $\alpha_i \in \mathbb{F}_{\text{fin}}$. Thus,

$$T(x) = \alpha_1 T(x_1) + \cdots + \alpha_n T(x_n) \approx \text{st}(\alpha_1)T(x_1) + \cdots + \text{st}(\alpha_n)T(x_n) \in W,$$

so $T(x) \in W_{\text{ns}} \subseteq W_{\text{fin}}$. □

We now introduce a very important class of linear transformations that will play a prominent role in the rest of these notes. In some sense, they are the transformations on infinite-dimensional spaces that behave most like transformations between finite-dimensional spaces.

Definition 10.20. If $T : V \to W$ is a linear transformation, we say that T is a compact transformation if T is a compact map of the associated metric spaces, that is, for every bounded $B \subseteq V$, there is a compact $K \subseteq W$ such that $T(B) \subseteq K$. (Or in nonstandard terms, $T(V_{\text{fin}}) \subseteq W_{\text{ns}}$.)

By Theorem 10.14, a compact linear transformation is automatically bounded.

Exercise 10.21. Let $\mathcal{B}_0(V, W)$ denote the set of compact linear transformations from V to W. Show that $\mathcal{B}_0(V, W)$ is a subspace of $\mathcal{B}(V, W)$.

Observe that the proof of Corollary 10.19 actually shows

Corollary 10.22. *If $T : V \to W$ is a linear transformation and V is finite-dimensional, then T is compact.*

In fact, we can generalize the previous corollary, but first we need a definition.

Definition 10.23. If $T : V \to W$ is a linear transformation, we say that T is *finite-rank* if $T(V)$ is a finite-dimensional subspace of W.

Certainly, if V is finite-dimensional, then $T : V \to W$ is of finite-rank.

Lemma 10.24. *If $T : V \to W$ is of finite rank and bounded, then T is compact.*

Proof. We know that $T(V_{\text{fin}}) \subseteq W_{\text{fin}}$. Set $W' := T(V)$. Then $W'_{\text{fin}} = W'_{\text{ns}}$ by Corollary 10.17. Thus $T(V_{\text{fin}}) \subseteq W'_{\text{fin}} = W'_{\text{ns}} \subseteq W_{\text{ns}}$, whence T is compact. □

Exercise 10.25. Let $\mathcal{B}_{00}(V, W)$ denote the set of finite-rank bounded transformations from V to W. Show that $\mathcal{B}_{00}(V, W)$ is a subspace of $\mathcal{B}_0(V, W)$.

We now consider convergence of transformation.

Lemma 10.26. *Suppose that $B \subseteq V$ is a bounded set and $T_n, T : V \to W$ are all bounded linear transformations. Further assume that $\|T_n - T\| \to 0$ as $n \to \infty$. Then $T_n \to T$ uniformly on B.*

Proof. Choose $M \in \mathbb{R}^{>0}$ such that $\|x\| \leq M$ for all $x \in B$. We need to show that $T_N(x) \approx T(x)$ for all $x \in B^*$ and all $N \in \mathbb{N}^* \setminus \mathbb{N}$. Well, $\|T_N(x) - T(x)\| \leq \|T_N - T\|\|x\| \leq M \cdot \|T_N - T\| \approx 0$ since $\|T_N - T\| \approx 0$. □

Exercise 10.27. Suppose that W is a Banach space. Prove that the normed space $\mathcal{B}(V, W)$ is also a Banach space.

Proposition 10.28. *Suppose that W is a Banach space, $T_n, T : V \to W$ are bounded linear operators and each T_n is compact. Further suppose that $\|T_n - T\| \to 0$ as $n \to \infty$. Then T is compact.*

Proof. Let $x \in V_{\text{fin}}$; we need $T(x) \in W_{\text{ns}}$. Let $B \subseteq V$ be a ball around 0 such that $x \in B^*$. Since T_n converges uniformly to T on B, we know that $T|B : B \to W$ is compact. Thus, $T(B)$ is contained in a compact subset of W, whence $T(B)^* \subseteq W_{\text{ns}}$. In particular, $T(x) \in W_{\text{ns}}$. □

The previous proposition can be rephased as saying that, when W is a Banach space, that $\mathcal{B}_0(V, W)$ is a *closed* subspace of $\mathcal{B}(V, W)$, whence also a Banach space. By Exercise 10.25, the closure of $\mathcal{B}_{00}(V, W)$ is a closed subspace of $\mathcal{B}_0(V, W)$, in symbols: $\overline{\mathcal{B}_{00}(V, W)} \subseteq \mathcal{B}_0(V, W)$. In the next section, we will encounter a certain class of Banach spaces where $\overline{\mathcal{B}_{00}(V, W)} = \mathcal{B}_0(V, W)$.

10.4. **Problems.** Throughout, V denotes a normed space. $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.

Problem 10.1. Suppose that $x \in V^*$ is such that $x \approx 0$. Show that there is $N \in \mathbb{N}^* \setminus \mathbb{N}$ such that $Nx \approx 0$.

Problem 10.2. Suppose that V is a Banach space and $(x_n \mid n \in \mathbb{N})$ is a sequence from V such that $\sum_{n=1}^{\infty} \|x_n\| < \infty$. Show that $\sum_{n=1}^{\infty} x_n$ converges in V.

For the next problem, you will need to use the following:

Fact 10.29. *If V is a normed space, then $\{x \in V \mid \|x\| \leq 1\}$ is compact if and only if V is finite-dimensional.*

Problem 10.3. Suppose that V is a Banach space and $T : V \to V$ is a compact linear operator.

(1) Show that the identity operator $I : V \to V$ is compact if and only if V is finite-dimensional.
(2) Suppose that $U : V \to V$ is any bounded linear operator. Show that $T \circ U$ and $U \circ T$ are also compact.
(3) Suppose that T is invertible. Show that T^{-1} is compact if and only if V is finite-dimensional.

Problem 10.4. Suppose $K : [0,1] \times [0,1] \to \mathbb{R}$ is a continuous function. Suppose that $T : C([0,1],\mathbb{R}) \to C([0,1],\mathbb{R})$ is defined by

$$T(f)(s) := \int_0^1 f(t)K(s,t)dt.$$

Show that T is a compact linear operator. (Hint: Use our earlier characterization of $C([0,1],\mathbb{R})_{\mathrm{ns}}$.) Such an operator is called a *Fredholm Integral Operator.*

Problem 10.5. Suppose that $\|\cdot\|_1$ and $\|\cdot\|_2$ are both norms on a vector space W. We say that $\|\cdot\|_1$ and $\|\cdot\|_2$ are *equivalent* if there exist constants $c, d \in \mathbb{R}^{>0}$ such that, for all $x \in W$, we have

$$c\|x\|_1 \leq \|x\|_2 \leq d\|x\|_1.$$

For $x, y \in W^*$ and $i = 1, 2$, let us write $x \approx_i y$ to mean $\|x - y\|_i \approx 0$.

(1) Suppose that, for all $x \in W^*$, if $x \approx_1 0$, then $x \approx_2 0$. Show that $\{\|x\|_2 \mid x \in W, \|x\|_1 \leq 1\}$ is bounded.
(2) Suppose that $\{\|x\|_2 \mid x \in W, \|x\|_1 \leq 1\}$ is bounded. Let
$$d := \sup\{\|x\|_2 \mid x \in W, \|x\|_1 \leq 1\}.$$
Show that $\|x\|_2 \leq d\|x\|_1$ for all $x \in W$.
(3) Show that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent iff for all $x \in W^*$, we have $x \approx_1 0$ iff $x \approx_2 0$.
(4) Show that $(W, \|\cdot\|_1)$ is a Banach space iff $(W, \|\cdot\|_2)$ is a Banach space.
(5) Suppose $A \subseteq W$. For $i = 1, 2$, say that A is *open$_i$* if A is open with respect to the metric associated to $\|\cdot\|_i$. Show that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent iff for all $A \subseteq W$, we have A is open$_1$ iff A is open$_2$. (In fancy language, this exercise says that two norms are equivalent if and only if they induce the same *topology* on W.)

Problem 10.6.

(1) Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$ (not necessarily the usual norm on $\mathbb{R}^n$). Suppose $x \in (\mathbb{R}^*)^n$. Show that $\|x\| \approx 0$ iff $|x_i| \approx 0$ for $i = 1, \ldots, n$.

(2) Conclude that any two norms on a finite-dimensional vector space are equivalent.
(3) For $f \in C([0,1],\mathbb{R})$, define $\|f\|_1 := \int_0^1 |f(x)|dx$. Show that $\|\cdot\|_1$ is a norm on $C([0,1],\mathbb{R})$.
(4) Show that $(C([0,1],\mathbb{R}), \|\cdot\|_1)$ is not a Banach space. Thus $\|\cdot\|_1$ is not equivalent to the norm $\|\cdot\|_\infty$ on $C([0,1],\mathbb{R})$ considered earlier. Hence, for infinite-dimensional vector spaces, there can exist inequivalent norms.
(5) Give a direct proof that the identity function

$$I : (C([0,1],\mathbb{R}), \|\cdot\|_1) \to (C([0,1],\mathbb{R}), \|\cdot\|_\infty)$$

is an unbounded linear operator. (This follows from earlier problems, but it would be nice to find $f \in C([0,1],\mathbb{R})^*$ such that $\|f\|_1 \in \mathbb{R}_{\text{fin}}$ but $\|f\|_\infty \notin \mathbb{R}_{\text{fin}}$.)

11. Hilbert Spaces

Once again, $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.

11.1. Inner product spaces.

Definition 11.1. An *inner product space* (i.p.s.) is a vector space V over $\mathbb{F}$ equipped with an *inner product*, which is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ satisfying, for all $x, y, z \in V$ and $\alpha \in \mathbb{F}$:

- $\langle x, x \rangle \in \mathbb{R}$ and $\langle x, x \rangle \geq 0$;
- $\langle x, x \rangle = 0 \Leftrightarrow x = 0$;
- $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (complex conjugate);
- $\langle \alpha x + y, z \rangle = \alpha \langle x, z \rangle + \langle y, z \rangle$.

Throughout, $(V, \langle \cdot, \cdot \rangle)$ denotes an inner product space over $\mathbb{F}$.

Exercise 11.2. Show that, for all $x, y, z \in V$ and $\alpha \in \mathbb{F}$:

(1) $\langle x, \alpha y + z \rangle = \overline{\alpha} \langle x, y \rangle + \langle x, z \rangle$.
(2) $\langle x, 0 \rangle = 0$.

Exercise 11.3. Define $\|\cdot\| : V \to \mathbb{R}$ by $\|x\| := \sqrt{\langle x, x \rangle}$. Show that $\|\cdot\|$ is a norm on V.

By the previous exercise, we can consider an i.p.s. over $\mathbb{F}$ as a normed space over $\mathbb{F}$, and hence as a metric space as well. We will often consider V as an i.p.s., normed space, and metric space all at the same time. (In fact, it is the interplay between these three structures on V that is what is most interesting.)

The following is encountered in a first course in linear algebra:

Theorem 11.4 (Cauchy-Schwarz Inequality)**.** *For all $x, y \in V$, we have $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$.*

Corollary 11.5. *Equip $V \times V$ with the metric d given by*

$$d((x_1, x_2), (y_1, y_2)) := \max\{\|x_1 - y_1\|, \|x_2 - y_2\|\}.$$

Then the inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ is continuous.

Proof. Fix $x, y, u, v \in V$. Then

$$\begin{aligned} \|\langle x, y\rangle - \langle u, v\rangle\| &= \|\langle x, y\rangle - \langle x, v\rangle + \langle x, v\rangle - \langle u, v\rangle\| \\ &\leq \|\langle x, y - v\rangle\| + \|\langle x - u, v\rangle\| \\ &\leq \|x\| \cdot \|y - v\| + \|x - u\| \cdot \|v\|. \end{aligned}$$

Now suppose that $(x, y) \in V \times V$ and $(u, v) \in \mu((x, y))$. Then $x \approx u$ and $y \approx v$. Since $\|x\|, \|v\| \in \mathbb{R}_{\text{fin}}$, it follows from the transfer of the above inequality that $\langle x, y\rangle \approx \langle u, v\rangle$. $\square$

Example 11.6. The main example of an i.p.s. encountered in a first course on linear algebra is $\mathbb{F}^n$, equipped with the *standard inner product* $\langle \vec{x}, \vec{y}\rangle := \sum_{i=1}^n x_i \overline{y_i}$. Observe that this induces the norm on $\mathbb{F}^n$ introduced in the previous section.

Example 11.7. The infinite-dimensional analogue of the previous example is ℓ^2 as defined in the previous section. The inner product on ℓ^2 is given by $\langle (x_n), (y_n)\rangle := \sum_{n=1}^\infty x_n \overline{y_n}$. To see that this sum converges, use Cauchy-Schwarz for $\mathbb{F}^m$ (applied to $(|x_1|, \ldots, |x_m|)$ and $(|y_1|, \ldots, |y_m|)$) to see that

$$\sum_{n=1}^m |x_n y_n| \leq \sqrt{\sum_{n=1}^m |x_n|^2 \cdot \sum_{n=1}^m |y_n|^2} \leq \sqrt{\sum_{n=1}^\infty |x_n|^2 \cdot \sum_{n=1}^\infty |y_n|^2}.$$

Now let $m \to \infty$. It is now easy to verify that the axioms for an i.p.s. hold. We should remark that of all the ℓ^p spaces, ℓ^2 is the only one that carries the structure of an i.p.s. and the above inner product on ℓ^2 induces the norm on ℓ^2 introduced in the previous section.

Example 11.8. Let $V = C([0, 1], \mathbb{F})$. Then V becomes an i.p.s. when equipped with the inner product given by $\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}dx$. How does the norm on V induced by the inner product compare with the norm placed on V in the previous section?

Definition 11.9. V is called a *Hilbert space* if the metric associated to V is complete.

In other words, an i.p.s. is a Hilbert space if the associated normed space is a Banach space.

Exercise 11.10.

(1) Prove that the inner product spaces in Exercises 11.6 and 11.7 are Hilbert spaces. (Your proof for ℓ^2 should probably be standard as we have yet to characterize ℓ^2_{ns}.)
(2) Prove that the inner product space in Exercise 11.8 is *not* a Hilbert space.

11.2. **Orthonormal bases and ℓ^2.**

Definition 11.11.

(1) If $x, y \in V$, we say that x and y are *perpendicular* or *orthogonal* if $\langle x, y\rangle = 0$. We sometimes write $x \perp y$ to indicate that x and y are orthogonal.

(2) A set $\{e_i \mid i \in I\}$ from V is called *orthonormal* if
- $\|e_i\| = 1$ for each $i \in I$, and
- e_i, e_j are orthogonal for all $i \neq j$.

Lemma 11.12. *If $\{e_i \mid i \in I\}$ is orthonormal, then it is also linearly independent.*

Proof. Suppose $\sum_{n=1}^{m} c_n e_{i_n} = 0$. Then

$$c_k = \sum_{n=1}^{m} c_n \langle e_{i_n}, e_{i_k}\rangle = \langle \sum_{n=1}^{m} c_n e_{i_n}, e_{i_k}\rangle = \langle 0, e_{i_k}\rangle = 0.$$

$\square$

Definition 11.13. An *orthonormal basis for V* is a maximal orthonormal sequence of vectors for V.

By Zorn's lemma, every inner product space has an orthonormal basis. One must be careful with the word basis here: while in finite-dimensional inner product spaces, an orthonormal basis is a basis (in the usual linear algebra sense), for infinite-dimensional inner product spaces, an orthonormal basis is never a basis. (In this setting, the usual notion of "basis" is called "Hamel basis" to help make the distinction.)

Fact 11.14. *Let $(e_n \mid n \in \mathbb{N})$ be an orthonormal set of vectors for the Hilbert space H. Then the following are equivalent:*

(1) *(e_n) is an orthonormal basis for V;*
(2) *If $v \in V$ is such that $v \perp e_n$ for each n, then $v = 0$;*
(3) *For all $v \in V$, there is a sequence (α_n) from $\mathbb{F}$ such that $\sum_{n=0}^{m} \alpha_n e_n$ converges to v as $m \to \infty$.*

You will prove the previous fact in the exercises. In (3) of the previous fact, we write $v = \sum_{n=0}^{\infty} \alpha_n e_n$.

Example 11.15. For $n \in \mathbb{N}$, let $b_n \in \ell^2$ be defined by

$$b_n := (0, 0, \ldots, 0, 1, 0, \ldots, 0),$$

where the 1 is in the n^{th} spot. Then certainly (b_n) is an orthonormal sequence. Fix $a = (a_1, a_2, \ldots) \in \ell^2$. Then

$$\|a - \sum_{n=1}^{m} a_n b_n\|^2 = \sum_{n=m+1}^{\infty} \|a_n\|^2 \to 0$$

as $m \to \infty$ since $a \in \ell^2$. It follows that (b_n) is an orthonormal basis for ℓ^2, often referred to as the *standard orthonormal basis for* ℓ^2. Observe that (b_n) is not a Hamel basis as the vector $(\frac{1}{n^2})$ is not in the span of (b_n).

Fact 11.16. *Every orthonormal basis for ℓ^2 is countable.*

More generally, for any inner product space V, every two orthonormal bases for V have the same cardinality. We will often refer to an inner product space as being *separable* when its orthonormal bases are countable. (It is a fact that all separable Hilbert spaces are isomorphic to ℓ^2, but we will not need this fact.)

Until further notice, let us fix a separable Hilbert space H (which you may think of as ℓ^2). We also fix an orthonormal basis $(e_n \mid n \in \mathbb{N})$ for H.

Lemma 11.17. *For $a = \sum_{n=0}^{\infty} \alpha_n e_n$ and $b = \sum_{n=0}^{\infty} \beta_n e_n$, we have $\langle a, b\rangle = \sum_{n=0}^{\infty} \alpha_n \overline{\beta_n}$.*

Proof. Fix $m \in \mathbb{N}$. Then it is straightforward to verify that

$$\langle \sum_{n=0}^{m} \alpha_n e_n, \sum_{n=0}^{m} \beta_n e_n\rangle = \sum_{n=0}^{m} \alpha_n \overline{\beta_n}.$$

It remains to let $m \to \infty$ and use the fact that the inner product is continuous. $\square$

Corollary 11.18. *For $a = \sum_{n=0}^{\infty} \alpha_n e_n$, we have $\|a\|^2 = \sum_{n=0}^{\infty} |\alpha_n|^2$.*

Corollary 11.19. *For $a \in H$, there is a unique sequence (α_n) from $\mathbb{F}$ such that $a = \sum_{n=0}^{\infty} \alpha_n e_n$.*

Proof. Suppose that $\sum_{n=0}^{\infty} \alpha_n e_n = a = \sum_{n=0}^{\infty} \alpha_n' e_n$. Then $\sum_{n=0}^{\infty} (\alpha_n - \alpha_n') e_n = 0$, whence $\sum_{n=0}^{\infty} |\alpha_n - \alpha_n'|^2 = \|0\|^2 = 0$. $\square$

Let's bring the nonstandard description into the picture: We get a nonstandard extension of the orthonormal basis $(e_n \mid n \in \mathbb{N}^*)$; this set is still orthonormal by transfer. Also, for every $a \in H^*$, there is a unique sequence $(\alpha_n \mid n \in \mathbb{N}^*)$ from $\mathbb{F}^*$ such that $a = \sum_{n \in \mathbb{N}^*} \alpha_n e_n$ in the sense that, for every $\epsilon \in (\mathbb{R}^{>0})^*$, there is $m_0 \in \mathbb{N}^*$ such that, for all $m \in \mathbb{N}^*$ with $m \geq m_0$, we have $\|a - \sum_{n=0}^{m} \alpha_n e_n\| < \epsilon$. But what do we mean by $\sum_{n=1}^{m} \alpha_n e_n$ if $m > \mathbb{N}$? Well, we have the set $\mathrm{FinSeq}(V)$ of finite sequences from V, which is a subset of $\mathcal{P}(\mathbb{N} \times V)$, and the corresponding function $\Sigma : \mathrm{FinSeq}(V) \to V$ given by adding the elements of the finite sequence. Consequently, we get a function $\Sigma : \mathrm{FinSeq}(V)^* \to V^*$, given by "adding" the elements of the *hyperfinite sequence.*

In a similar vein, we have $\|a\|^2 = \sum_{n \in \mathbb{N}^*} |\alpha_n|^2$, with the interpretation of the sum as in the previous paragraph. We have thus proven:

Proposition 11.20. *For $a = \sum_{n \in \mathbb{N}^*} \alpha_n e_n \in H^*$, we have $a \in H_{\text{fin}}$ if and only if $\sum_{n \in \mathbb{N}^*} |\alpha_n|^2 \in \mathbb{R}_{\text{fin}}$.*

Characterizing the nearstandard points is a bit more subtle:

Theorem 11.21. *Suppose that* $a = \sum_{n\in\mathbb{N}^*} \alpha_n e_n$ *is in* H_{fin}. *Then* $a \in H_{\text{ns}}$ *if and only if, for every* $k > \mathbb{N}$, *we have* $\sum_{n>k} |\alpha_n|^2 \approx 0$. *In this case,* $\text{st}(a) = \sum_{n\in\mathbb{N}} \text{st}(\alpha_n) e_n$.

Proof. First suppose that $a \in H$, so $a = \sum_{n\in\mathbb{N}} \alpha_n e_n$. For $k \in \mathbb{N}$, set $r_k = \sum_{n=k+1}^{\infty} |\alpha_n|^2$. Thus, $r_k \to 0$ as $k \to \infty$ and hence $r_k \approx 0$ if $k > \mathbb{N}$; this proves the result for $a \in H$. Now suppose that $a \in H_{\text{ns}}$, say $a \approx b \in H$. Write $b = \sum_{n=1}^{\infty} \beta_n e_n$. Fix $k > \mathbb{N}$. Then, by the transfer of the triangle inequality for ℓ^2, we have

$$\sqrt{\sum_{n>k} |\alpha_n|^2} \leq \sqrt{\sum_{n>k} |\alpha_n - \beta_n|^2} + \sqrt{\sum_{n>k} |\beta_n|^2} = (\dagger) + (\dagger\dagger).$$

Since $a \approx b$, we have that $\|a - b\|^2 = \sum_{n\in\mathbb{N}^*} |\alpha_n - \beta_n|^2 \approx 0$, whence $(\dagger) \approx 0$. It remains to prove that $(\dagger\dagger) \approx 0$. However, since $b \in H$, by the first part of the proof, we know that $(\dagger\dagger) \approx 0$.

Now suppose that $\sum_{n>k} |\alpha_n|^2 \approx 0$ for every $k > \mathbb{N}$; we must show that $a \in H_{\text{ns}}$. Since $a \in H_{\text{fin}}$, we know that $\sum_{n\in\mathbb{N}^*} |\alpha_n|^2 \in \mathbb{R}_{\text{fin}}$; say $\sum_{n\in\mathbb{N}^*} |\alpha_n|^2 \leq M \in \mathbb{R}^{>0}$. In particular, $\alpha_n \in \mathbb{F}_{\text{fin}}$ for each $n \in \mathbb{N}^*$. For $n \in \mathbb{N}$, set $\beta_n := \text{st}(\alpha_n)$. We claim that $b := \sum_{n\in\mathbb{N}} \beta_n e_n$ defines an element of H. For $m \in \mathbb{N}$, define $s_m := \sum_{n=0}^{m} |\beta_n|^2$. Then $s_m \approx \sum_{n=0}^{m} |\alpha_n|^2 \leq M$. Hence, (s_m) is a bounded, nondecreasing sequence in $\mathbb{R}$, whence convergent. For $m, k \in \mathbb{N}$ with $m \leq k$, we have

$$\|\sum_{n=0}^{m} \beta_n e_n - \sum_{n=0}^{k} \beta_n e_n\|^2 \leq \sum_{n=m+1}^{k} |\beta_n|^2.$$

Since (s_m) is Cauchy, it follows that the partial sums of $\sum_{n=0}^{\infty} \beta_n e_n$ are Cauchy, whence, by completeness, $\sum_{n=0}^{\infty} \beta_n e_n$ defines an element of H.

It remains to verify that $a \approx b$. To see this, observe that, for any $k \in \mathbb{N}^*$, we have

$$\|a - b\|^2 = \sum_{n\in\mathbb{N}^*} |\alpha_n - \beta_n|^2 = \sum_{n\leq k} |\alpha_n - \beta_n|^2 + \sum_{n>k} |\alpha_n - \beta_n|^2.$$

If $k \in \mathbb{N}$, then $\sum_{n\leq k} |\alpha_n - \beta_n|^2 \approx 0$; thus, by the Infinitesimal Prolongation Theorem, there is $k > \mathbb{N}$ such that $\sum_{n\leq k} |\alpha_n - \beta_n|^2 \approx 0$. On the other hand,

$$\sum_{n>k} |\alpha_n - \beta_n|^2 \leq \sum_{n>k} |\alpha_n|^2 + \sum_{n>k} |\beta_n|^2.$$

By assumption, $\sum_{n>k} |\alpha_n|^2 \approx 0$, whilst $\sum_{n>k} |\beta_n|^2 \approx 0$ by the forward direction of the theorem and the fact that b is standard. Consequently, $\|a - b\|^2 \approx 0$. $\square$

For $a = \sum_{n=0}^{\infty} \alpha_n e_n \in H$ and $m \in \mathbb{N}$, set $P(m, a) := P_m(a) = \sum_{n=0}^{m} \alpha_n e_n \in H$. We thus have maps $P : \mathbb{N} \times H \to H$ and, for $n \in \mathbb{N}$, $P_n : H \to H$.

Exercise 11.22. For $n \in \mathbb{N}$, P_n is a bounded linear transformation with $\|P_n\| = 1$.

Extending to the nonstandard universe, we get a map $P : \mathbb{N}^* \times H^* \to H^*$. For $N \in \mathbb{N}^*$, we define $P_N : H^* \to H^*$ by $P_N(a) = P(N, a)$. For $a = \sum_{n\in\mathbb{N}^*} \alpha_n e_n \in H^*$, we may view $P_N(a)$ as the hyperfinite sum $\sum_{n=0}^{N} \alpha_n e_n \in H^*$.

Theorem 11.23. *Suppose $N > \mathbb{N}$ and $a \in H_{\text{ns}}$. Then $P_N(a) \approx a$.*

Proof. Write $a = \sum_{n\in\mathbb{N}^*} \alpha_n e_n$. Then, by Theorem 11.21, we have

$$\|a - \sum_{n=0}^{N} \alpha_n e_n\|^2 = \|\sum_{n>N} \alpha_n e_n\|^2 = \sum_{n>N} |\alpha_n|^2 \approx 0.$$

□

Recall that we have a norm $\|\cdot\|$ on $\mathcal{B}(H)$. Going to the nonstandard universe, we get an internal norm $\|\cdot\| : \mathcal{B}(H)^* \to \mathbb{R}^*$. By transfer, for $T \in \mathcal{B}(H)^*$, we have $\|T\|$ is the internal suprememum of the set $\{\|T(x)\| \mid x \in H^*,\ \|x\| = 1\}$. By the transfer of Exercise 11.22, for $N > \mathbb{N}$, we have $P_N \in \mathcal{B}(H)^*$ and $\|P_N\| = 1$.

Corollary 11.24. *Suppose that $T : H \to H$ is a compact operator and $N > \mathbb{N}$. Then $\|T - P_N T\| \approx 0$.*

Proof. Suppose $a \in H^*$ and $\|a\| = 1$. Then $a \in H_{\text{fin}}$, whence $T(a) \in H_{\text{ns}}$ by the compactness of T. Thus, by Theorem 11.23, $T(a) \approx P_N T(a)$. Since the internal supremum of an internal set of infinitesimals is once again infinitesimal, this shows that $\|T - P_N T\| \approx 0$. □

The following lemma will be crucial in the proof of the Bernstein-Robinson Theorem in Section 14.

Lemma 11.25. *Suppose that $T \in \mathcal{B}(H)$ is compact, $\{e_1, e_2, \ldots\}$ is an orthonormal basis for H, and $[a_{jk}]$ is the "matrix" for T with respect to this basis, that is, for all $k \geq 1$, we have $T(e_k) = \sum_{j\in\mathbb{N}} a_{jk} e_j$. Then $a_{jk} \approx 0$ for all $j \in \mathbb{N}^* \setminus \mathbb{N}$ and $k \in (\mathbb{N}^*)^{>0}$. (So the "infinite" rows of the matrix for T consist entirely of infinitesimals.)*

Proof. Fix $k \in (\mathbb{N}^*)^{>0}$. Since T is compact and e_k is finite, $T(e_k) \in H_{\text{ns}}$, say $T(e_k) \approx y$. Write $y = \sum_{i\in\mathbb{N}^*} y_i e_i$. Set

$$\eta := \|T(e_k) - y\|^2 = \|\sum_{i\in\mathbb{N}^*} (a_{ik} - y_i) e_i\|^2 = \sum_{i\in\mathbb{N}^*} |a_{ik} - y_i|^2.$$

Since $T(e_k) \approx y$, we have $\eta \approx 0$. Now suppose $j \in \mathbb{N}^* \setminus \mathbb{N}$. Then $|a_{jk} - y_j|^2 \leq \eta$, whence $|a_{jk} - y_j| \approx 0$. Since $y \in H$, $y_j \approx 0$. Since $|a_{jk}| \leq |a_{jk} - y_j| + |y_j|$, we get that $a_{jk} \approx 0$. □

11.3. **Orthogonal projections.** In this subsection, we assume that H is an arbitrary Hilbert space. For $u, v \in H$, we have the *Parallelogram Identity*:

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

If, in addition, u and v are orthogonal, we have the *Pythagorean Theorem*: $\|u+v\|^2 = \|u\|^2 + \|v\|^2$. By induction, we see that whenever $u_1, \ldots, u_n$ are mutually orthogonal, then $\|u_1 + \cdots + u_n\|^2 = \|u_1\|^2 + \cdots + \|u_n\|^2$.

Lemma 11.26. *If $\{t_1, \ldots, t_k\}$ is an orthonormal set in H, then $\operatorname{sp}(t_1, \ldots, t_k)$ is a closed subspace of H.*

Later we will be able to remove the "orthonormal" assumption.

Proof. Set $E := \operatorname{sp}(t_1, \ldots, t_k)$. Suppose $x \in E^*$ and $y \in H$ are such that $x \approx y$; we need $y \in E$. By transfer, there are $\alpha_1, \ldots, \alpha_k \in \mathbb{F}^*$ such that $x = \alpha_1 t_1 + \cdots + \alpha_k t_k$. By the (transfer of the) Pythagorean theorem, we see that $\|x\|^2 = \|\alpha_1 t_1\|^2 + \cdots + \|\alpha_k t_k\|^2 = |\alpha_1|^2 + \cdots + |\alpha_k|^2$. Since $x \in H_{\text{ns}}$, we see that $\|x\| \in \mathbb{R}_{\text{fin}}$, whence each $\alpha_i \in \mathbb{F}_{\text{fin}}$. Set $z := \operatorname{st}(\alpha_1)t_1 + \cdots + \operatorname{st}(\alpha_k)t_k \in E$. By the Pythagorean theorem again, we see that

$$\|x - z\|^2 = |\alpha_1 - \operatorname{st}(\alpha_1)|^2 + \cdots + |\alpha_k - \operatorname{st}(\alpha_k)|^2 \approx 0.$$

By uniqueness of standard part, we see that $y = z \in E$. $\square$

The next result is crucial in the study of Hilbert spaces.

Theorem 11.27 (Existence of Orthogonal Projection)**.** *Suppose that E is a closed subspace of H. Then for each $x \in H$, there is a unique $y \in E$ such that $\|x - y\| \leq \|x - z\|$ for all $z \in E$.*

Proof. The result is obvious if $x \in E$ (take $y = x$). Thus, we may assume that $x \notin E$. Set $\alpha := \inf\{\|x - z\| \mid z \in E\}$; note that $\alpha > 0$ since E is closed and $x \notin E$. Fix $\epsilon \in \mu^{>0}$. By transfer, there is $z \in E^*$ such that $\alpha \leq \|x - z\| < \alpha + \epsilon$; in particular, $\|x - z\| \approx \alpha$.

Claim: $z \in E_{\text{ns}}$.
Proof of Claim: Suppose, towards a contradiction, $z \notin E_{\text{ns}}$. Since H is complete and E is closed in H, we have that E is complete. Consequently, since $z \notin E_{\text{ns}}$, we have that $z \notin E_{\text{pns}}$, whence there is $r \in \mathbb{R}^{>0}$ such that $\|z - w\| \geq r$ for all $w \in E$. Since $\alpha < \sqrt{\alpha^2 + \frac{r^2}{4}}$, we have $w \in E$ such that $\|x - w\| < \sqrt{\alpha^2 + \frac{r^2}{4}}$. By the Parallelogram Identity, we have:

$$\|(x - w) + (x - z)\|^2 + \|w - z\|^2 = 2(\|x - w\|^2 + \|x - z\|^2).$$

Now $\|(x - w) + (x - z)\| = 2\|x - \frac{1}{2}(w + z)\| \geq 2\alpha$ by transfer and the fact that $\frac{1}{2}(w + z) \in E^*$. Since $\|x - z\| \approx \alpha$, we have that $2\|x - z\|^2 < 2\alpha^2 + \frac{r^2}{2}$.

Thus:

$$\begin{aligned}\|w-z\|^2 &= 2(\|x-w\|^2+\|x-z\|^2)-\|(x-w)+(x-z)\|^2\\ &< 2(\alpha^2+\frac{r^2}{4})+2\|x-z\|^2-4\alpha^2\\ &< 2(\alpha^2+\frac{r^2}{4})+2\alpha^2+\frac{r^2}{2}-4\alpha^2\\ &< r^2.\end{aligned}$$

This contradicts the fact that $\|z-w\|\geq r$.

Set $y:=\mathrm{st}(z)\in E$. Since $x-z\approx x-y$, we see that $\|x-y\|=\alpha$, whence $\|x-y\|\leq\|x-z\|$ for all $z\in E$.

To prove uniqueness: suppose that $y'\in E$ also satisfies the conclusion of the theorem. In particular, $\|x-y\|=\|x-y'\|=\alpha$. Then

$$\|y-y'\|^2=2\|x-y\|^2+2\|x-y'\|^2-4\|x-\frac{1}{2}(y+y')\|^2\leq 4\alpha^2-4\alpha^2=0,$$

whence $y=y'$. □

Definition 11.28. If E is a closed subspace of H and $x\in E$, we let $P_E(x)$ denote the unique element of y fulfilling the conclusion of the previous theorem; we refer to $P_E(x)$ as the *orthogonal projection of x onto E.*

Definition 11.29. For E a closed subspace of H, we set

$$E^\perp:=\{x\in H\mid x\perp z\text{ for all }z\in E\}.$$

Until further notice, E denotes a closed subspace of H.

Lemma 11.30. *$E^\perp$ is a closed subspace of H.*

Proof. We leave it to the reader to check that $E^\perp$ is a subspace of H. To check that $E^\perp$ is closed, suppose that $x\in(E^\perp)^*$ and $y\in H$ are such that $x\approx y$; we need to show that $y\in E^\perp$. Fix $z\in E$; we need $\langle y,z\rangle=0$. Well, by continuity of the inner product, $\langle y,z\rangle\approx\langle x,z\rangle=0$ by transfer. Since $\langle y,z\rangle\in\mathbb{R}$, it follows that $\langle y,z\rangle=0$. □

Much of what follows is easy and standard but will be needed in the next sections. We will thus omit many proofs.

Lemma 11.31. *For any $x\in H$, $x-P_E(x)\in E^\perp$.*

Lemma 11.32. *For $x,y\in H$, we have $y=P_E(x)$ if and only if $y\in E$ and $x-y\in E^\perp$. Thus, if $x\in E^\perp$, then $P_E(x)=0$.*

Theorem 11.33 (Gram-Schmidt Process)**.** *Let $(u_n\mid n\in I)$ be a sequence of linearly independent vectors in H, where $I=\{1,\ldots,m\}$ for some $m\in\mathbb{N}$ or $I=\mathbb{N}$. Then there is an orthonormal sequence $(w_n\mid n\in I)$ such that, for each $k\in I$, $\mathrm{sp}(u_1,\ldots,u_k)=\mathrm{sp}(t_1,\ldots,t_k)$.*

Proof. Set $w_1 := \frac{u_1}{\|u_1\|}$. Inductively, suppose that $w_1, \ldots, w_k$ have been constructed. Set $v_{k+1} := u_{k+1} - P_E(u_{k+1})$, where $E := \mathrm{sp}(w_1, \ldots, w_k)$, a closed subspace of H by Lemma 11.26. Set $w_{k+1} := \frac{v_{k+1}}{\|v_{k+1}\|}$. Then $w_{k+1} \in E^\perp$, whence $\{w_1, \ldots, w_{k+1}\}$ is orthonormal. □

Corollary 11.34. *Any finite-dimensional subspace of H is closed.*

Proof. By the Gram-Schmidt process, any finite-dimensional subspace of H has an orthonormal basis that is also a Hamel basis; thus, the subspace is closed by Lemma 11.26. □

Exercise 11.35. Suppose that $u_1, \ldots, u_k$ are orthonormal and $E = \mathrm{sp}(u_1, \ldots, u_k)$. Then, for all $x \in H$, we have $P_E(x) = \sum_{i=1}^k \langle x, u_i \rangle u_i$.

Let us momentarily return to the discussion of the previous subsection. Suppose that H is separable and that $(e_n \mid n \geq 1)$ is an orthonormal basis for H. Set $P_m : H \to H$ as before, namely $P_m(a) = \sum_{n=1}^m \alpha_n e_n$, where $a = \sum_{n=1}^\infty \alpha_n e_n$. Set $E_m := \mathrm{sp}(e_1, \ldots, e_m)$, a closed subspace of H. Then, for $a = \sum_{n=1}^\infty \alpha_n e_n$, we have:

$$\begin{aligned}
P_{E_m}(a) &= \sum_{n=1}^m \langle a, e_n \rangle e_n \\
&= \sum_{n=1}^m \langle \lim_{k \to \infty} (\sum_{j=1}^k \alpha_j e_j), e_n \rangle e_n \\
&= \sum_{n=1}^m \lim_{k \to \infty} \langle \sum_{j=1}^k \alpha_j e_j, e_n \rangle e_n \\
&= \lim_{k \to \infty} \sum_{n=1}^m \sum_{j=1}^k \alpha_j \langle e_j, e_n \rangle e_n \\
&= \sum_{j=1}^m \alpha_j e_j = P_m(a).
\end{aligned}$$

Lemma 11.36. *P_E is a bounded linear operator. If $E \neq \{0\}$, then $\|P_E\| = 1$.*

Lemma 11.37. *Suppose that F is also a closed subspace of H and $E \perp F$, that is, $x \perp y$ for all $x \in E$ and $y \in F$. Set*

$$G := E + F := \{x + y \mid x \in E, y \in F\}.$$

Then G is a closed subspace of H and for all $z \in G$, there are unique *$x \in E$ and $y \in F$ such that $z = x + y$.*

In the situation of the previous lemma, we write $G = E \oplus F$, the *direct sum of E and F*.

Lemma 11.38. $H = E \oplus E^\perp$.

Lemma 11.39. $(E^{\perp})^{\perp} = E$.

Lemma 11.40. *Suppose that* $G = E \oplus F$. *Then* $P_G = P_E + P_F$. *Consequently,* $I = P_E + P_{E^{\perp}}$.

Now suppose that E_1, E_2, E_3 are closed subspaces such that $E_1 \perp E_2$, $E_2 \perp E_3$ and $E_1 \perp E_3$. Then $E_1 \perp (E_2 \oplus E_3)$, $(E_1 \oplus E_2) \perp E_3$, and $E_1 \oplus (E_2 \oplus E_3) = (E_1 \oplus E_2) \oplus E_3$. We may thus unambiguously write $E_1 \oplus E_2 \oplus E_3$. Ditto for any finite number of mutually perpendicular closed subspaces $E_1, \ldots, E_n$ of H; we often write the direct sum in the compact notation $\bigoplus_{i=1}^n E_i$.

Corollary 11.41. *If* $G = \bigoplus_{i=1}^n E_i$, *then* $P_G = P_{E_1} + \cdots + P_{E_n}$.

Now suppose that we have countably many mutually perpendicular subspaces $(E_n \mid n \geq 1)$. How should we define their direct sum? First, two lemmas, the second of which we give a short nonstandard proof.

Lemma 11.42. *Suppose that* $(G_n \mid n \geq 1)$ *is a sequence of subspaces of* H *with* $G_n \subseteq G_{n+1}$ *for all* $n \geq 1$. *Then* $G = \bigcup_{n\geq 1} G_n$ *is a subspace of* H.

Lemma 11.43. *Suppose that* G *is a subspace of* H. *Then* $\overline{G}$ *is also a subspace of* H.

Proof. Suppose that $x, y \in \overline{G}$; we show that $x + y \in \overline{G}$. (The proof for closure under scalar multiplication is similar.) Since $x \in \overline{G}$, there is $x' \in G^*$ such that $x \approx x'$. Similarly, there is $y' \in G^*$ such that $y \approx y'$. Thus, $x + y \approx x' + y' \in G^*$ (by transfer), whence $x + y \in \overline{G}$. $\square$

Returning to the situation preceding the lemmas, suppose that $(E_n \mid n \geq 1)$ are mutually perpendicular closed subspaces of H. Set $G_n := \bigoplus_{i=1}^n E_i$, a closed subspace of H. We then define $\bigoplus_{n=1}^\infty E_i := \overline{\bigcup_{n\geq 1} G_n}$, a closed subspace of H by the previous two lemmas.

Lemma 11.44. *Suppose that* $(G_n \mid n \geq 1)$ *is a sequence of subspaces of* H *with* $G_n \subseteq G_{n+1}$ *for all* $n \geq 1$ *and* $G = \overline{\bigcup_{n\geq 1} G_n}$. *Then, for all* $x \in H$, *we have* $P_{G_n}(x) \to P_G(x)$ *as* $n \to \infty$.

Corollary 11.45. *Suppose that* $E = \bigoplus_{i=1}^\infty E_i$. *Then, for all* $x \in H$, *we have* $\sum_{i=1}^n P_{E_i}(x) \to P_E(x)$ *as* $n \to \infty$.

11.4. **Hyperfinite-dimensional subspaces.** Once again, H denotes an arbitrary Hilbert space. Let $\mathcal{E}$ denote the set of finite-dimensional subspaces of H, whence $\mathcal{E} \subseteq \mathcal{P}(H)$. We will refer to elements of $\mathcal{E}^*$ as *hyperfinite-dimensional subspaces of* H^*. We have a map $P : \mathcal{E} \times H \to H$ given by $P(E, x) := P_E(x)$. We thus get a nonstandard extension $P : \mathcal{E}^* \times H^* \to H^*$, whence it makes sense to speak of the orthogonal projection map $P_E : H^* \to H^*$ for $E \in \mathcal{E}^*$. Similary, we have a map $\dim : \mathcal{E} \to \mathbb{N}$, whence we get a map $\dim : \mathcal{E}^* \to \mathbb{N}^*$.

If H is separable and $(e_n \mid n \geq 1)$ is an orthonormal basis for H, then by transfer, for $N \in \mathbb{N}^*$, we have $H_N := \mathrm{sp}(e_1, \ldots, e_N) \in \mathcal{E}^*$, the *internal* *subspace of H^* spanned by* $e_1, \ldots, e_N$. The elements of H_N are internal linear combinations (with coefficients in $\mathbb{F}^*$) of $e_1, \ldots, e_N$. By transfer, $\dim(H_N) = N$.

The following theorem is crucial in applications of nonstandard methods to the study of infinite-dimensional linear algebra:

Theorem 11.46. *If H is an infinite-dimensional Hilbert space, then there is $E \in \mathcal{E}^*$ such that $H \subseteq E$.*

Proof. Exercise. □

Recall from Problem 9.5 the notion of the standard part of an internal set.

Exercise 11.47. Suppose that $E \in \mathcal{E}^*$ and $x \in H$. Then $x \in \mathrm{st}(E)$ if and only if $P_E x \approx x$.

Lemma 11.48. *If $E \in \mathcal{E}^*$, then $\mathrm{st}(E)$ is a closed linear subspace of H.*

Proof. $\mathrm{st}(E)$ is closed by Problem 9.5. It is straightforward to verify that $\mathrm{st}(E)$ is a subspace of H. □

11.5. **Problems.**

Problem 11.1. Suppose $(\lambda_n \mid n \geq 1)$ is a sequence from $\mathbb{C}$. Consider the function $D : \mathbb{C}^{\mathbb{N}} \to \mathbb{C}^{\mathbb{N}}$ given by $D((x_n)) := (\lambda_n x_n)$.

(1) Show that $D \in B(\ell^2)$ if and only if the sequence (λ_n) is bounded. (We are sort of abusing notation and letting D denote its restriction to ℓ^2.)
(2) Show that $D \in B_\circ(\ell^2)$ if and only if $\lim_{n\to\infty} \lambda_n = 0$.

Problem 11.2. Consider the function $T : \ell^2 \to \ell^2$ given by

$$T(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots).$$

Show that $T \in B(\ell^2)$ and $T \notin B_\circ(\ell^2)$. What is $\|T\|$? (T is called the *left shift operator*.)

Problem 11.3. Suppose that T is a compact operator on the separable Hilbert space H. Let $\epsilon \in \mathbb{R}^{>0}$. Show that there is a bounded finite-rank linear operator T' on ℓ^2 such that $\|T - T'\| < \epsilon$. In fancier language, we are showing that the set of finite-rank linear operators on ℓ^2, $B_{\circ\circ}(\ell^2)$, is dense in the set of compact linear operators on ℓ^2, $B_\circ(\ell^2)$. (Hint: The desired T' will be of the form $P_k T$ for some $k \in \mathbb{N}$. Consider $k > \mathbb{N}$ and use underflow.)

Problem 11.4. Suppose that H is a separable Hilbert space and $x \in H_{\mathrm{ns}}$. Suppose that (e_n) is an orthonormal basis for ℓ^2 and $x = \sum_{n\in\mathbb{N}^*} x_n e_n$. Further suppose that $x_n \approx 0$ for all $n \in \mathbb{N}$. Show that $x \approx 0$.

Problem 11.5. Suppose $P : H \to H$ is an idempotent bounded operator on a Hilbert space H. Show that P is compact if and only if P is a finite-rank operator.

Problem 11.6. Suppose that E is any linear subspace of any unitary space U. Show that $E^{\perp}$ is a linear subspace of U.

We will need the following notation for the next problem. If $w = a + bi \in \mathbb{C}$, then $\mathrm{Re}(w) := a$ and $\mathrm{Im}(w) := b$.

Problem 11.7. Suppose that E is a closed linear subspace of the Hilbert space H. Suppose $x \notin E$. We aim to show that $x - P_E(x) \in E^{\perp}$. Set $\alpha := \|x - P_E(x)\| > 0$. Fix $z \in E \setminus \{0\}$.

(1) Fix $\lambda \in \mathbb{R} \setminus \{0\}$. Show that $\lambda^2\|z\|^2 - 2\lambda \,\mathrm{Re}(\langle x - P_E(x), z\rangle) > 0$. (Hint: Start with $\alpha^2 < \|x - (P_E(x) + \lambda z)\|^2$.)
(2) Considering
$$\lambda := \frac{\mathrm{Re}(\langle x - P_E(x), z\rangle)}{\|z\|^2},$$
conclude that $\mathrm{Re}(\langle x - P_E(x), z\rangle) = 0$.
(3) Show that $\langle x - P_E(x), z\rangle = 0$, and thus $x - P_E(x) \in E^{\perp}$.

Problem 11.8. Suppose $(e_1, \ldots, e_k)$ is an orthonormal sequence of vectors in the Hilbert space H. Suppose $E = \mathrm{span}(e_1, \ldots, e_k)$. Show that, for all $v \in H$, we have
$$P_E(x) = \sum_{i=1}^{k} \langle v, e_i\rangle e_i.$$

Problem 11.9. Suppose that (e_n) is an orthonormal sequence from the Hilbert space H. Fix $v \in H$. Prove *Bessel's inequality*:
$$\sum_{n=1}^{\infty} |\langle v, e_n\rangle|^2 \leq \|v\|^2.$$
(Hint: Use the last problem and the fact that a projection P_n has norm 1.)

Problem 11.10. Suppose that (e_n) is an orthonormal sequence from the Hilbert space H. Show that the following are equivalent:

(1) (e_n) is an orthonormal basis for ℓ^2;
(2) For every $v \in \ell^2$, if $\langle v, e_n\rangle = 0$ for all $n \geq 1$, then $v = 0$.
(3) For every $v \in \ell^2$, there is a sequence (α_n) such that
$$v = \sum_{n=1}^{\infty} \alpha_n e_n.$$

Problem 11.11. Suppose that H is a finite-dimensional Hilbert space. Show that an orthonormal basis for H is also a Hamel basis.

Problem 11.12. Show that the orthonormal basis (b_n) for ℓ^2 defined in lecture is not a Hamel basis.

It can be shown that any orthonormal basis for an infinite dimensional Hilbert space is not a Hamel basis. We outline the proof of this here for ℓ^2. We first need a definition. If X is a metric space and $A \subseteq X$, we say that

A is *nowhere dense* if $(\overline{A})^\circ = \emptyset$; in English, if the interior of the closure of A is empty.

Problem 11.13. Suppose that H is a Hilbert space and $(e_n \mid n \geq 1)$ is a linearly independent set of vectors. For $n \geq 1$, set

$$E_n := \operatorname{span}(e_1, \ldots, e_n).$$

Show that each E_n is a nowhere dense subset of V. (Hint: We already know that each E_n is closed. (Why?) It remains to show that each E_n has empty interior. Use the nonstandard characterization of the interior to do this.)

The *Baire Category Theorem* states that in a complete metric space, the union of a countable number of nowhere dense sets is nowhere dense.

Problem 11.14. Use the Baire category theorem and the previous problem to show that a Hamel basis of an infinite dimensional Hilbert space must be uncountable. Conclude that no orthonormal basis for ℓ^2 can be a Hamel basis.

In working with Hilbert spaces, orthonormal bases are more useful than Hamel bases. One can show that any two orthonormal bases for a Hilbert space have the same cardinality, called the *dimension* of the Hilbert space. One can then show that two Hilbert spaces are isomorphic (that is, there is a bijective linear transformation between them preserving the inner product) if and only if they have the same dimension. This is not true if we use dimension to mean the cardinality of a Hamel basis (which is well-defined). Indeed, for $s \in \mathbb{R}$, let $u_s : \mathbb{R} \to \mathbb{C}$ be defined by $u_s(t) := e^{ist}$. Let V be the vector space consisting of all finite linear combinations of the u_s. For $f, g \in H$, one can show that

$$\langle f, g \rangle := \lim_{N \to \infty} \frac{1}{2N} \int_{-N}^{N} f(t)\overline{g(t)}dt$$

exists and defines an inner product on V. Let H denote the completion of V. Then H is naturally a Hilbert space. One can show that ℓ^2 and H both have Hamel dimension $2^{\aleph_0}$. However, $(u_s \mid s \in \mathbb{R})$ is an orthornormal basis for H, so H has dimension $2^{\aleph_0}$ and cannot be isomorphic to ℓ^2 as a Hilbert space.

12. Weekend Problem Set #2

Problem 12.1. Assume the nonstandard extension is $\aleph_1$-saturated. Let A be an internal set and B a (not necessarily internal) subset of A. We say that B is Σ^0_1 if there is a sequence $(C_k \mid k \in \mathbb{N})$ of internal subsets of A such that $B = \bigcup_k C_k$. We say that B is Π^0_1 if there exists a sequence $(C_k \mid k \in \mathbb{N})$ of internal subsets of A such that $B = \bigcap_k C_k$.

(1) Show that $\mathbb{R}_{\text{fin}}$ is a Σ^0_1 subset of $\mathbb{R}^*$ and μ is a Π^0_1 subset of $\mathbb{R}^*$. (This is why Σ^0_1 sets are sometimes called *galaxies* and Π^0_1 sets are sometimes called *monads*.)

(2) Suppose $(C_k \mid k \in \mathbb{N})$ is a sequence of internal subsets of an internal set A. Let $B = \bigcup_k C_k$. Suppose that B is internal. Show that there is $n \in \mathbb{N}$ such that $B = C_1 \cup \cdots \cup C_n$. (Hint: Reduce to the case that $C_n \subseteq C_{n+1}$ for all $n \in \mathbb{N}$. Use the Countable Comprehension Principle and Underflow.)
(3) Suppose $(C_k \mid k \in \mathbb{N})$ is a sequence of internal subsets of an internal set A. Let $B = \bigcap_k C_k$. Suppose that B is internal. Show that there is $n \in \mathbb{N}$ such that $B = C_1 \cap \cdots \cap C_n$.
(4) Suppose that B_1 is a Σ^0_1 subset of the internal set A and B_2 is a Π^0_1 subset of A. Suppose that $B_1 \subseteq B_2$. Show that there is an internal subset E of A such that $B_1 \subseteq E \subseteq B_2$. Conclude that any subset of A which is both Σ^0_1 and Π^0_1 must be internal. (This is a handy way of showing that certain sets are internal.)

We should remark that (2) from the previous exercise is the key observation behind the *Loeb measure* concept, which will be discussed later in these notes. This concept is one of the central ideas in applications of nonstandard analysis to probability theory, stochastic analysis, and mathematical finance (to name a few).

There is an ultrapower construction for our more general nonstandard framework. If $\mathcal{U}$ is a nonprincipal ultrafilter on $\mathbb{N}$, then one can set $X^* := X^{\mathcal{U}}$, where $X^{\mathcal{U}}$ is defined just as $\mathbb{R}^{\mathcal{U}}$. This construction has the usual advantage of being "concrete." In particular, we can identify the internal sets in a concrete fashion as follows.

Suppose that we have a family $(X_n \mid n \in \mathbb{N})$ of sets. Then we can define their *ultraproduct* $\prod_{\mathcal{U}} X_n$ to be $\prod_n X_n / \sim_{\mathcal{U}}$. (Thus, the ultrapower is the special case of the ultraproduct where all X_n's are equal.)

Problem 12.2. Prove that $A \subseteq X^{\mathcal{U}}$ is internal if and only if there are subsets $A_n \subseteq X$ for $n \in \mathbb{N}$ such that $A = \prod_{\mathcal{U}} A_n$. (So the internal subsets of $X^{\mathcal{U}}$ are the ones that are almost everywhere given coordinatewise by subsets of X.)

Problem 12.3. Show that $A \subseteq X^{\mathcal{U}}$ is hyperfinite if and only if there are *finite* subsets $A_n \subseteq X$ such that $A = \prod_{\mathcal{U}} A_n$. Use this fact to explain how to define the internal cardinality $|A|$ of a hyperfinite subset A of $X^{\mathcal{U}}$.

Problem 12.4. Prove that a nonstandard extension constructed using a nonprincipal ultrafilter $\mathcal{U}$ on $\mathbb{N}$ is automatically $\aleph_1$-saturated. (Hint: this is essentially a diagonal argument.)

Problem 12.5. (Assume as much saturation as you like) Suppose that G is a group. Then G is said to be *nilpotent* if there exists a finite sequence of subgroups $G_1, \ldots, G_n$ of G such that

$$\{1\} \leq G_1 \leq G_2 \leq \ldots \leq G_n = G$$

and such that for all $i \in \{1, \ldots, n\}$, all $g \in G$, and all $h \in G_i$, we have $ghg^{-1}h^{-1} \in G_{i-1}$. If n is the smallest length of such a sequence of subgroups for G, we call n the *nilpotency class* of G.

(1) G is nilpotent of nilpotency class 1 if and only if G is nontrivial (i.e. $G \neq \{1\}$) and abelian (i.e. $xy = yx$ for all $x, y \in G$.)
(2) Show that G^* is a group, as is any subset of G^* closed under the extension of the group operations.
(3) Suppose $A \subseteq G$. The *subgroup of G generated by A*, denoted $\langle A \rangle$, is
$$\langle A \rangle := \bigcap \{H \mid H \leq G \text{ and } A \subseteq H\}.$$
A subgroup H of G is called *finitely generated* if $H = \langle A \rangle$ for some finite $A \subseteq G$. Discuss how the nonstandard extension of the set of finitely generated subgroups of G is the set of hyperfinitely generated subgroups of G^*.
(4) Show that there is a hyperfinitely generated subgroup H of G^* such that $G \leq H$.
(5) Suppose that G is *locally nilpotent*, that is, suppose that every finitely generated subgroup of G is nilpotent. Further suppose that the nilpotency class of any finitely generated subgroup of G is less than or equal to n. Show that G is nilpotent of nilpotency class less than or equal to n. (You will have to use the fact that a subgroup of a nilpotent group of nilpotency class $\leq n$ is itself a nilpotent group of nilpotency class $\leq n$. Try proving this fact!)
(6) Why doesn't your proof in (5) work if G is locally nilpotent with unbounded nilpotency class? (There are some difficult open problems about locally nilpotent groups of unbounded nilpotency class that some group theorists hope might be solved by nonstandard methods.)

Problem 12.6. (Assume $\aleph_1$-saturation) Suppose that X is a normed space over $\mathbb{F}$ and $E \subseteq X^*$ is an *internal subspace*, that is, E is internal, $0 \in E$, E is closed under addition, and E is closed under multiplication by elements of $\mathbb{F}^*$. For example, $E := X^*$ is an internal subspace of X^*. Define
$$E_{\text{fin}} := \{x \in E \mid \|x\| \in \mathbb{F}_{\text{fin}}\}$$
and
$$\mu_E := \{x \in E \mid \|x\| \in \mu(0)\}.$$

(1) Show that E_{fin} is a vector space over $\mathbb{F}$ and that μ_E is a subspace of E_{fin}. We set $\hat{E} := E_{\text{fin}}/\mu_E$ and call it the *nonstandard hull of E*. For $x \in E_{\text{fin}}$, we often write $\hat{x}$ instead of $x + \mu_E$.
(2) For $\hat{x} \in \hat{E}$, define $\|\hat{x}\| := \text{st}(\|x\|)$. Show that this definition is independent of the coset representative and that $\|\cdot\| : \hat{E} \to \mathbb{R}$ is a norm on $\hat{E}$.
(3) Show that $\hat{E}$ is a *Banach* space. (Notice as a consequence that even if X was incomplete, $\widehat{X^*}$ is automatically complete.)

(4) Show that there is a map $\iota : \hat{E} \to \widehat{X^*}$ given by $\iota(x + \mu_E) = x + \mu_{X^*}$. Show that ι is a norm-preserving linear map whose image is closed in $\hat{X}^*$. (This exercise allows us to treat $\hat{E}$ as a Banach subspace of $\widehat{X^*}$.)

The nonstandard hull construction is extremely useful in applications of nonstandard methods to functional analysis. Of particular interest is when E is a hyperfinite-dimensional subspace of X^* containing X, for then one can often apply facts about finite-dimensional normed spaces to $\hat{E}$ and then have these facts "trickle down" to X.

13. The Spectral Theorem for compact hermitian operators

In this section, H denotes an arbitrary Hilbert space over $\mathbb{C}$.

Definition 13.1. If $T : H \to H$ is linear, then $\lambda \in \mathbb{C}$ is called an *eigenvalue for T* if there is a nonzero $x \in H$ such that $Tx = \lambda x$; such an x will be called an *eigenvector for T corresponding to λ*.

If λ is an eigenvalue for T, then $\{x \in H \mid Tx = \lambda x\}$ is called the *eigenspace of T corresponding to λ*; as the name indicates, it is a subspace of H. It is also easy to verify that eigenspaces are *closed* in H.

Definition 13.2. $T : H \to H$ is *Hermitian* if $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$.

Exercise 13.3. Suppose that E is a closed subspace of H. Prove that P_E is a Hermitian operator.

The following theorem is standard fare for a first undergraduate course in linear algebra:

Theorem 13.4 (Spectral Theorem for Hermitian Operators on Finite-Dimensional Hilbert Spaces)**.** *Suppose $T : H \to H$ is a Hermitian operator and H is finite-dimensional. Then there is an orthonormal basis for H consisting of eigenvectors of T. Moreover, if $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ are the distinct eigenvalues of T with corresponding eigenspaces $W_1, \ldots, W_k$, then $W_i \perp W_j$ for all $i \neq j$, $H = W_1 \oplus \cdots \oplus W_k$, and $T = \lambda_1 P_{W_1} + \cdots + \lambda_k P_{W_k}$, the so-called* spectral resolution *of T.*

Proof. (Sketch) Set $n := \dim(H)$.

Step 1: Show that T has an eigenvalue λ_1 with corresponding eigenvector x_1; we may further assume $\|x_1\| = 1$. (This part does not use the fact that T is Hermitian, but only that we are working over $\mathbb{C}$ so the characteristic polynomial of T splits.)

Step 2: Let $W := \mathrm{sp}(x_1)$. If $H = W$, we are done. Otherwise, $W^\perp$ is a Hilbert space of dimension $n-1$. Furthermore, since T is Hermitian, one can easily verify that $T(W^\perp) \subseteq W^\perp$, whence we can consider the Hermitian operator $T|W^\perp : W^\perp \to W^\perp$. By induction, $W^\perp$ has an orthonormal basis $x_2, \ldots, x_n$ consiting of eigenvectors of T. Since $W \perp W^{\perp}$, $\{x_1, \ldots, x_n\}$ is an orthonormal basis for H consisting of eigenvectors of T.

Step 3: Show that each λ_i is a real number. We leave this as an exercise for the reader.

Step 4: Show that $W_i \perp W_j$ for $i \neq j$.

Step 5: Since we have a basis of eigenvectors, $H = W_1 \oplus \cdots \oplus W_k$.

Step 6: Just compute

$$T = TI = T(P_{W_1} + \cdots + P_{W_k}) = TP_{W_1} + \cdots + TP_{W_k} = \lambda_1 P_{W_1} + \cdots + \lambda_k P_{W_k}.$$

□

What about Hermitian operators on infinite-dimensional Hilbert spaces? We immediately run into problems:

Exercise 13.5. Let $H = C([0,1], \mathbb{C})$ with the inner product

$$\langle f, g \rangle := \int_0^1 f(t)\overline{g(t)}dt.$$

Define $T : H \to H$ by $T(f) = t \cdot f$. Show that T is a bounded operator on H that is not compact and does not have any eigenvalues.

In order to obtain eigenvalues, we will need to study *compact* Hermitian operators. For such operators, there is a suitable version of the Spectral Theorem that is a rather straightforward generalization of the finite-dimensional case. In fact, we will use the method of hyperfinite approximation and the transfer of the finite-dimensional spectral theorem to obtain the spectral theorem for compact Hermitian operators. We should mention that there are a plethora of spectral theorems for other sorts of operators on Hilbert spaces, but the functional analysis needed to study them (e.g. spectral measures) is beyond the scope of these notes.

For the rest of this section, we assume that H is an infinite-dimensional Hilbert space. Moreover, we fix a hyperfinite-dimensional subspace E of H^* (that is, $E \in \mathcal{E}^*$) such that $H \subseteq E$; see Theorem 11.46. We let P denote P_E.

Lemma 13.6. *If $x \in H_{\text{ns}}$, then $Px \approx x$.*

Proof. Write $x = y + \epsilon$, where $y \in H$ and $\epsilon \approx 0$. Then $Px = Py + P\epsilon$; since $\|P\| = 1$, $P\epsilon \approx 0$. Since $y \in H \subseteq E$, $Py = y$. Thus, $Px \approx Py = y \approx x$. □

If $T \in \mathcal{B}(H)$, we define $T' : E \to E$ by $T'(x) = PT(x)$. It is straightforward to check that $\|T'\| \leq \|T\|$ and $T'x = Tx$ for $x \in H$. A slightly less trivial observation is:

Lemma 13.7. *If T is Hermitian, so is T'.*

Proof. For $x, y \in E$, we calculate:

$$\langle T'x, y \rangle = \langle Tx, Py \rangle = \langle Tx, y \rangle = \langle x, Ty \rangle = \langle Px, Ty \rangle = \langle x, PTy \rangle = \langle x, T'y \rangle.$$

□

Lemma 13.8. *If $x \in H_{\rm ns}$, then $T'Px \approx Tx$.*

Proof. By Lemma 13.6, $Px \approx x$. Since T is continuous, we have $TPx \approx Tx$. Since $Tx \in H_{\rm ns}$ (again, by continuity of T), we have $PTx \approx Tx$. Thus, $T'Px = PTPx \approx PTx \approx Tx$. $\square$

For the rest of this section, we assume that $T \in \mathcal{B}(H)$ is a nonzero, *compact* Hermitian operator and $T' : E \to E$ is defined as above. In this case, T' is *internally compact* in the following precise sense:

Lemma 13.9. *If $x \in E \cap H_{\rm fin}$, then $T'x \in H_{\rm ns}$.*

Proof. Since T is compact, $Tx \in H_{\rm ns}$. By Lemma 13.6, $T'x = PTx \approx Tx$, whence $T'x \in H_{\rm ns}$ as well. $\square$

Exercise 13.10. Suppose that $\lambda \in \mathbb{C}^*$ is an eigenvalue of T'. Show that $|\lambda| \leq \|T'\|$, whence $\lambda \in \mathbb{C}_{\rm fin}$.

Lemma 13.11. *Suppose that $\lambda \in \mathbb{C}^*$ is an eigenvalue of T' with corresponding eigenvector $x \in E$ of norm 1. Further suppose that $\lambda \not\approx 0$. Then $x \in H_{\rm ns}$ and $\operatorname{st}(x)$ is an eigenvector of T of norm 1 corresponding to the eigenvalue $\operatorname{st}(\lambda)$.*

Proof. By the internal compactness of T', we know that $T'x \in H_{\rm ns}$. But $T'x = \lambda x$, so since $\frac{1}{\lambda} \in \mathbb{C}_{\rm fin}$, we have $x = \frac{1}{\lambda}T'x \in H_{\rm ns}$. Also, by Lemma 13.8, we have

$$T(\operatorname{st}(x)) \approx Tx \approx T'Px = T'x = \lambda x \approx \operatorname{st}(\lambda)\operatorname{st}(x).$$

Consequently, $T(\operatorname{st}(x)) = \operatorname{st}(\lambda)\operatorname{st}(x)$. It remains to observe that $\operatorname{st}(x)$ has norm 1; however, this follows from the fact that x has norm 1 and $\|\operatorname{st}(x)\|$ is a real number. $\square$

At this point, we apply the (transfer of the) finite-dimensional Spectral Theorem to T'; in particular, we get a hyperfinite sequence of eigenvalues $\lambda_1, \ldots, \lambda_\nu$ (perhaps with repetitions), where $\nu := \dim(E)$. Without loss of generality (by reordering if necessary), we may assume that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_\nu|.$$

We let $\{x_1, \ldots, x_\nu\}$ be an orthonormal basis for E, where x_i is an eigenvalue for T' corresponding to λ_i.

We now worry about how many repetitions we might have amongst the eigenvalues. We show that a noninfinitesimal eigenvalues can only be repeated a finite number of times. First, we need:

Lemma 13.12. *Suppose $\epsilon \in \mathbb{R}^{>0}$ and $|\lambda_k| \geq \epsilon$. Then $k \in \mathbb{N}$.*

Proof. By the Pythagorean Theorem, $\|x_i - x_j\| = \sqrt{2}$ for $i, j = 1, \ldots, k$. We define the internal sequence $(s_i \ : \ i \in \mathbb{N}^*)$ by $s_i = x_i$ for $i = 1, \ldots, k$, while $s_i = 0$ for $i > k$. Suppose, towards a contradiction, that $k > \mathbb{N}$. Then $\|s_i - s_j\| = \sqrt{2}$ for distinct $i, j \in \mathbb{N}$, whence, by Theorem 9.36, $s_i = x_i \notin H_{\rm ns}$ for some $i \in \mathbb{N}^*$, contradicting Lemma 13.11. $\square$

Corollary 13.13. *Suppose that $\lambda_{l+1} = \lambda_{l+2} = \cdots = \lambda_{l+k} \not\approx 0$. Then $k \in \mathbb{N}$.*

Proof. By the previous lemma, $l + k \in \mathbb{N}$, whence $k \in \mathbb{N}$. $\square$

We now remove repetitions: let $\kappa_1, \ldots, \kappa_\eta$ enumerate all the distinct eigenvalues of T', again ordered so that $|\kappa_1| \geq |\kappa_2| \geq \cdots \geq |\kappa_\eta|$. For $i = 1, \ldots, \eta$, we let $\{x^i_1, \ldots, x^i_{m_i}\}$ be an orthonormal basis for the eigenspace W_i corresponding to κ_i. By the previous corollary, we know that $m_i \in \mathbb{N}$ for each $i \in \{1, \ldots, \eta\}$ with $\kappa_i \not\approx 0$. Moreover, we have $E = W_1 \oplus \cdots \oplus W_\eta$, $P = P_{W_1} + \cdots + P_{W_\eta}$, and $T' = \kappa_1 P_{W_1} + \cdots + \kappa_\eta P_{W_\eta}$.

Which of the κ_i's are noninfinitesimal? To answer this question, we set $X := \{i \in \mathbb{N}^* \mid 1 \leq i \leq \eta \text{ and } \kappa_i \not\approx 0\}$.

Lemma 13.14. *$X \neq \emptyset$.*

Proof. Suppose, toward a contradiction, that $X = \emptyset$. Fix $b \in H$ such that $Tb \neq 0$, this is possible since $T \neq 0$. Write $b = b_1 + \cdots + b_\eta$, with $b_i \in W_i$ for each i. Then $Tb = T'b = T'b_1 + \cdots + T'b_\eta = \kappa_1 b_1 + \cdots + \kappa_\eta b_\eta$. However, by the Pythagorean Theorem,

$$\|\kappa_1 b_1 + \cdots + \kappa_\eta b_\eta\|^2 = \sum_{i=1}^{\eta} |\kappa_i|^2 \|b_i\|^2 \leq |\kappa_1|^2 \sum_{i=1}^{\eta} \|b_i\|^2 = |\kappa_1|^2 \|b\|^2 \approx 0,$$

a contradiction. $\square$

Lemma 13.15. *$X \subseteq \mathbb{N}$.*

Proof. Fix $n \in X$ and choose $\epsilon > 0$ such that $|\kappa_n| \geq \epsilon$. Recall that $\kappa_n = \lambda_m$ for some $m \in \{1, \ldots, \nu\}$. By Lemma 13.12, $m \in \mathbb{N}$. Since $n \leq m$, we have $n \in \mathbb{N}$. $\square$

We now know that one of the two situations occurs: either $X = \mathbb{N}^{>0}$ or $X = \{1, \ldots, k\}$ for some $k \in \mathbb{N}$. Both situations are in fact possible and must be treated slightly differently.

We set $y^j_i := \mathrm{st}(x^j_i)$ for $j \in X$ and $i \in \{1, \ldots, m_j\}$. If $i \neq i'$, then by continuity of the inner product, we know that $\langle y^j_i, y^j_{i'} \rangle \approx \langle x^j_i, x^j_{i'} \rangle = 0$, whence the y^j_i's are pairwise perpendicular. We set $\tilde{W}_j := \mathrm{sp}(y^j_1, \ldots, y^j_{m_j})$, so $\dim(\tilde{W}_j) = m_j$. In an analogous fashion, one shows that $\tilde{W}_j \perp \tilde{W}_{j'}$ for distinct $j, j' \in X$.

Lemma 13.16. *For $x \in H$ and $j \in X$, we have $P_{\tilde{W}_j}(x) \approx P_{W_j}(x)$.*

Proof.

$$P_{\tilde{W}_j}(x) = \sum_{i=1}^{m_j} \langle x, y^j_i \rangle y^j_i \approx \sum_{i=1}^{m_j} \langle x, x^j_i \rangle x^j_i = P_{W_j}(x).$$

$\square$

We now set $K := \bigoplus_{j \in X} \tilde{W}_j$; this is either a finite or infinite direct sum, depending on whether X is finite or infinite. We set $\tilde{W}_0 := K^\perp$.

For simplicity, set $\alpha_j := \mathrm{st}(\kappa_j)$. We can now state the Spectral Theorem in case W is finite.

Theorem 13.17 (Spectral Theorem- Case 1). *If $X = \{1, \ldots, k\}$ for some $k \in \mathbb{N}$, then the eigenvalues of T are $0, \alpha_1, \ldots, \alpha_k$ and*

$$T = \alpha_1 P_{\tilde{W}_1} + \cdots + \alpha_k T_{\tilde{W}_k}$$

is the spectral resolution of T.

Proof. Fix $x \in H$; then $T(x) = T'(x)$. Consequently,

$$T(x) = T'(x) \sum_{i=1}^{\eta} \kappa_j P_{W_j} = \sum_{j=1}^{k} \kappa_j P_{W_j}(x) + \sum_{j=k+1}^{\eta} \kappa_j P_{W_j}(x) = (\dagger) + (\dagger\dagger).$$

By the previous lemma, $(\dagger) \approx \sum_{j=1}^{k} \alpha_j P_{\tilde{W}_j}(x) \in H$. We need to show that $(\dagger\dagger) \approx 0$. This again follows from a Pythagorean Theorem calculation, as $\|(\dagger\dagger)\|^2 \leq |\kappa_{k+1}|^2 \|x\|^2 \approx 0$. This proves the Spectral Resolution part of the Theorem. We next show that 0 is an eigenvalue. First, since K is finite-dimensional while H is infinite-dimensional, we have $\tilde{W}_0 \neq \{0\}$. If $x \in \tilde{W}_0 \setminus \{0\}$, then by the Spectral Resolution, we have $T(x) = \sum_{j=1}^{k} \alpha_j P_{W_j}(x) = 0$, whence x is an eigenvector of T corresponding to 0. It remains to show that $\{0, \alpha_1, \ldots, \alpha_k\}$ are *all* of the eigenvalues of T; this will again follow from the Spectral Resolution of T. To see this, suppose that $\alpha \neq 0$ is an eigenvalue of T with corresponding eigenvector x. Then

$$\sum_{j=0}^{k} \alpha P_{\tilde{W}_j}(x) = \alpha x = Tx = \sum_{i=1}^{k} \alpha_i P_{\tilde{W}_i}(x).$$

Thus, $\alpha P_{\tilde{W}_0}(x) = 0$ and $\alpha P_{\tilde{W}_j}(x) = \alpha_j P_{\tilde{W}_j}(x)$ for $j = 1, \ldots, k$ (since elements of $\tilde{W}_0, \tilde{W}_1, \ldots, \tilde{W}_k$ are pairwise perpendicular and thus linearly independent). Since $x \neq 0$, there is $j \in \{1, \ldots, k\}$ such that $P_{\tilde{W}_j}(x) \neq 0$; for this j, we have $\alpha = \alpha_j$. □

In order to deal with the case that $X = \mathbb{N}^{>0}$, we need one final lemma.

Lemma 13.18. *If $X = \mathbb{N}^{>0}$, then $\lim_{j \to \infty} \alpha_j = 0$.*

Proof. Since $|\alpha_n|$ is nonincreasing and bounded below, it suffices to prove that 0 is a limit point of $|\alpha_n|$. We consider the extension of the sequence $(\alpha_n \mid n \in \mathbb{N}^{>0})$ to a sequence $(\alpha_n \mid n \in (\mathbb{N}^{>0})^*)$. By the Infinitesimal Prolongation Theorem, there is $N \in \mathbb{N}^* \setminus \mathbb{N}$, $N \leq \eta$ such that $\alpha_N \approx \kappa_N$. But $\kappa_N \approx 0$ since $N \notin X$. Thus, $\alpha_N \approx 0$ and hence 0 is a limit point of (α_n). □

We are now ready to state:

Theorem 13.19 (Spectral Theorem-Case 2). *If $X = \mathbb{N}^{>0}$, then the* nonzero *eigenvalues are $\alpha_1, \alpha_2, \ldots$ and $T = \sum_{j=1}^{\infty} \alpha_j P_{\tilde{W}_j}$ is the spectral resolution of T, that is, $T(x) = \sum_{i=1}^{\infty} \alpha_j P_{\tilde{W}_j}(x)$ for all $x \in H$.*

Proof. Fix $x \in H$. Set $r_n := \|T(x) - \sum_{j=1}^n \alpha_j P_{\tilde{W}_j}(x)\|$; we need to show that $r_n \to 0$ as $n \to \infty$. Again, we know that $\sum_{j=1}^n \alpha_j P_{\tilde{W}_j}(x) \approx \sum_{j=1}^n \kappa_j P_{W_j}(x)$ and $T(x) = T'(x) = \sum_{j=1}^\eta \kappa_j P_{W_j}(x)$. Thus, $r_n \approx \|\sum_{j=n+1}^\eta \kappa_j P_{W_j}(x)\|$ for all $n \in \mathbb{N}$. We play a Pythagorean game again:

$$\|\sum_{j=n+1}^{\eta} \kappa_j P_{W_j}(x)\|^2 = \sum_{j=n+1}^{\eta} |\kappa_j|^2 \|P_{W_j}(x)\|^2 \leq |\kappa_{n+1}|^2 \sum_{j=1}^{\eta} \|P_{W_j}(x)\|^2 = |\kappa_{n+1}|^2 \|x\|^2.$$

Thus $r_n \leq |\alpha_{n+1}|^2 \|x\|^2$ for all $n \in \mathbb{N}$. Since $\alpha_n \to 0$ as $n \to \infty$, we see that $r_n \to 0$ as $n \to \infty$. The claim about the nonzero eigenvalues of T being $\{\alpha_1, \alpha_2, \ldots,\}$ follows from the Spectral Resolution of T in a manner similar to Case 1 and is left as an exercise. □

13.1. **Problems.**

Problem 13.1. Suppose that T is a Hermitian operator on a unitary space V. Show that every eigenvalue of T is a real number.

Problem 13.2. Let V be the unitary space $C([0,1], \mathbb{C})$ endowed with the inner product

$$\langle f, g \rangle := \int_0^1 f(t)\overline{g(t)}dt.$$

Define $T : V \to V$ by $T(f) := tf$, i.e. $T(f)(t) := tf(t)$.

(1) Show that T is a bounded Hermitian operator on V.
(2) Show that T has no eigenvalues.
(3) Show that T is not compact. (Don't just say if it were compact, it would have eigenvalues by the Spectral Theorem. Show directly that T is not compact.)

Problem 13.3. For each of the following bounded linear operators T given below, explain why they are compact and Hermitian. Then find the eigenvalues, eigenspaces, and projections, yielding the spectral decomposition for T.

(1) Let A be the matrix
$$\begin{pmatrix} 3 & -4 \\ -4 & 3 \end{pmatrix}.$$
Define $T : \mathbb{C}^2 \to \mathbb{C}^2$ by $T(x) = Ax$.
(2) Define $T : \ell^2 \to \ell^2$ by $T((x_k \mid k \geq 1)) = (\frac{x_k}{k} \mid k \geq 1)$.

There is a Spectral Theorem for Hermitian operators which need not be compact. The statement of the spectral theorem is more involved (and a little less satisfying), so we will not give it here. However, the results in the remaining problems are ingredients towards proving this more general Spectral Theorem. The interested reader can find a complete discussion of this in Chapter 5, Section 5 of [1].

We let H be a Hilbert space and T a (not necessarily compact) Hermitian operator on H. E, P, T' and the sequence $(\lambda_i \mid i \leq \nu)$ are defined exactly as they are in the notes. We also set $J := [-\|T\|, \|T\|]$.

Problem 13.4. For $n \geq 0$, show that $(T')^n = \sum_{i=1}^{\nu} \lambda_i^n P_i$. Conclude that for any polynomial $p(z) \in \mathbb{R}^*[z]$, we have $p(T') = \sum_{i=1^\nu} p(\lambda_i)P_i$.

For any internal function $f : J^* \to \mathbb{R}^*$, set $f(T') = \sum_{i=1}^{\nu} f(\lambda_i)P_i$.

Problem 13.5. For any internal $f : J^* \to \mathbb{R}^*$, show that $f(T')$ is Hermitian.

Problem 13.6. Suppose $f, g : J^* \to \mathbb{R}^*$ are internal. Show that:
(1) $(f + g)(T') = f(T') + g(T')$;
(2) $(f \cdot g)(T') = f(T') \cdot g(T')$;
(3) $(cf)(T) = cf(T)$ for any $c \in \mathbb{R}^*$.

These properties are often referred to as the Operational Calculus.

Problem 13.7. Suppose that $f, g : J^* \to \mathbb{R}^*$ are internal functions and $f(c) \leq g(c)$ for all $c \in J^*$. Show that $f(T') \leq g(T')$.

We call an internally bounded linear operator (what does this mean?) U on E *nearstandard* if for every $x \in E \cap H_{\text{ns}}$, we have $U(x) \in H_{\text{ns}}$.

Problem 13.8. Suppose that U is nearstandard. Show that $\|U\| \in \mathbb{R}_{\text{fin}}$. (Careful: This is not immediate from the definition)

If U is nearstandard, define $\text{st}(U) : H \to H$ by $\text{st}(U)(x) = \text{st}(U(x))$ for all $x \in H$.

Problem 13.9. Suppose that U is nearstandard. Show that $\text{st}(U)$ is a bounded linear operator on H and $\|\text{st}(U)\| \leq \text{st}(\|U\|)$.

Problem 13.10. Show that, for any $n \in \mathbb{N}$, $(T')^n$ is nearstandard. Conclude that for any $p(z) \in \mathbb{R}[z]$, we have $p(T')$ is nearstandard.

Problem 13.11. Suppose that $f : J \to \mathbb{R}$ is continuous.
(1) Show that $f(T')$ is nearstandard. (Hint: Suppose, towards a contradiction, that $x \in E \cap H_{\text{ns}}$ is such that $f(T')(x) \notin H_{\text{ns}}$. Thus, there is $r \in \mathbb{R}^{>0}$ such that $\|f(T')(x) - b\| > r$ for all $b \in H$. By the Weierstrauss Approximation Theorem, there is a polynomial $p(z) \in \mathbb{R}[z]$ such that $|f(c) - p(c)| < r$ for all $c \in J$.)
(2) Show that $\text{st}((f(T')))$ is Hermitian.
(3) Suppose that $f(c) \geq 0$ for all $c \in J$. Show that $\text{st}(f(T')) \geq 0$, that is, $\langle \text{st}(f(T'))x, x\rangle \geq 0$ for all $x \in H$.

14. The Bernstein-Robinson Theorem

In this section, we will prove a theorem which was the first serious theorem proven using nonstandard methods. Before we can state this theorem, we need some background. Throughout this section, all normed spaces are over $\mathbb{C}$.

Definition 14.1. Suppose that X is a Banach space, $T \in \mathcal{B}(X)$, and $Y \subseteq X$ is a subspace. We say that Y is *T-invariant* if $T(Y) \subseteq Y$.

If X is a Banach space and $T \in \mathcal{B}(X)$, then clearly $\{0\}$ and X are T-invariant; these are said to be the *trivial* T-invariant subspaces. Clearly, if $\dim(X) = 1$, then X has no nontrivial T-invariant subspaces.

Theorem 14.2. *If X is a finite-dimensional Banach space with* $\dim(X) \geq 2$ *and $T \in \mathcal{B}(X)$, then X has a nontrivial T-invariant subspace.*

Proof. Set $n := \dim(X) \geq 2$. Fix $x \in X \setminus \{0\}$. Since $\{x, Tx, \ldots, T^n x\}$ has $n+1$ vectors, it must be linearly dependent, whence we can find scalars $\alpha_0, \ldots, \alpha_n \in \mathbb{C}$, not all 0, such that $\alpha_0 x + \alpha_1 Tx + \ldots + \alpha_n T^n x = 0$. Factor the polynomial $\alpha_0 + \alpha_1 z + \ldots + \alpha_n z^n = \alpha(z - \lambda_1) \cdots (z - \lambda_n)$, where $\alpha, \lambda_i \in \mathbb{C}$. Then $\alpha(T - \lambda_1 I) \cdots (T - \lambda_n I)x = 0$. It follows that there is $i \in \{1, \ldots, n\}$ such that $\ker(T - \lambda_i I) \neq \{0\}$. Observe that $\ker(T - \lambda_i I) = \{x \in X \mid T(x) = \lambda_i x\}$ is T-invariant. If $\ker(T - \lambda_i I) \neq X$, then it is a nontrivial T-invariant subspace. Otherwise, $Tx = \lambda_i x$ for all $x \in X$, whence it follows that $T = \lambda_i I$. But then, every subspace is T-invariant, so any 1-dimensional subspace will be a nontrivial T-invariant subspace. □

Thus, for "small" Banach spaces, non-trivial T-invariant subspaces always exist. At the opposite extreme, if X is a non-separable Banach space, then for any $x \in X \setminus \{0\}$, the closed linear span of $\{x, Tx, T^2 x, \ldots\}$ is a nontrival T-invariant subspace. What about separable, infinite-dimensional Banach spaces? Well, it was first shown by Enflo that there are quite exotic separable Banach spaces that possess bounded linear operators without any nontrivial, closed invariant subspaces. But what about the least exotic Banach spaces: Hilbert spaces? Well, the following is still open:

The Invariant Subspace Problem: Suppose that H is a separable, infinite-dimensional Hilbert space and $T \in \mathcal{B}(H)$. Does H have a nontrivial, closed T-invariant subspace?

In the rest of this section, H will denote a separable, infinite-dimensional Hilbert space.

Theorem 14.3. *If $T \in \mathcal{B}(H)$ is a compact, Hermitian operator, then H has a nontrivial, closed T-invariant subspace.*

Proof. By the Spectral Theorem, there is an eigenvalue λ for T. The associated eigenspace is a closed, T-invariant subspace that is not $\{0\}$. If the associated eigenspace were all of H, then $T = \lambda I$ and once again any 1-dimensional subspace is a nontrivial, closed T-invariant subspace. □

Notice that the proofs of the preceding theorems utilized eigenvalues in an essential way. Even so, the next example and exercises show that eigenvalues are not necessary for invariant subspaces.

Example 14.4. Let $S : \ell^2 \to \ell^2$ be the "right-shift" operator, that is, $S(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$. Notice $\|S\| = 1$, so $S \in \mathcal{B}(\ell^2)$. For $n > 0$, set $M_n := \{(x_1, x_2, \ldots) \in \ell^2 \mid x_k = 0 \text{ for } k = 1, 2, \ldots, n\}$. Then M_n is a nontrivial, closed S-invariant subspace.

Exercise 14.5. For S as in the previous example, show that S is not compact and has no eigenvalues.

Exercise 14.6. Let $W : \ell^2 \to \ell^2$ be given by

$$W(x_1, x_2, \ldots) = (0, x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \ldots).$$

Show that W is compact but does not have any eigenvalues. Once again, the various M_n are nontrivial W-invariant subspaces.

Before nonstandard methods, the state-of-the-art result was the following strengthening of Theorem 14.3:

Theorem 14.7 (von Neumann, 1930s)**.** *If $T \in \mathcal{B}(H)$ is compact, then H has a nontrivial, closed T-invariant subspace.*

Definition 14.8. $T \in \mathcal{B}(H)$ is said to be *polynomially compact* if there is a nonzero polynomial $p(z) \in \mathbb{C}[z]$ such that $p(T)$ is compact.

Our goal is to prove the following:

Theorem 14.9 (Bernstein-Robinson 1966)**.** *If $T \in \mathcal{B}(H)$ is polynomially compact, then H has a nontrivial, closed T-invariant subspace.*

To be fair, we should say that after the Bernstein-Robinson theorem, a more general theorem was proven whose proof is much more elementary (modulo some standard facts from a first graduate course in functional analysis):

Theorem 14.10 (Lomonosov, 1973)**.** *If $T \in \mathcal{B}(H)$ is such that there is a nonzero $K \in \mathcal{B}(H)$ such that K is compact and $TK = KT$, then H has a nontrivial, closed T-invariant subspace.*

(To see how Lomonosov implies Bernstein-Robinson, observe that T commutes with $p(T)$.)

The main idea behind the proof of the Bernstein-Robinson Theorem is fairly easy to explain. Once again, it is an instance of the method of Hyperfinite Approximation. Fix an orthonormal basis $\{e_1, e_2, \ldots\}$ for H and fix $\nu \in \mathbb{N}^* \setminus \mathbb{N}$. Set $H_\nu := \mathrm{sp}(e_1, \ldots, e_\nu)$. Let $P_\nu : H^* \to H_\nu$ be the orthogonal projection onto H_ν. Set $T_\nu : H^* \to H^*$ by $T_\nu = P_\nu T P_\nu$; observe that $T_\nu(H^*) \subseteq H_\nu$. We also set $\tilde{T} := T_\nu | H_\nu = (P_\nu T)|H_\nu$, so $\tilde{T}$ is an internally linear operator on H_ν. Observe also that

$$\|\tilde{T}\| \leq \|T_\nu\| = \|P_\nu T P_\nu\| \leq \|P_\nu\| \|T\| \|P_\nu\| = \|T\|.$$

We recall the following standard fact from linear algebra (whose proof we include for the sake of completeness).

Theorem 14.11 (Existence of Upper Triangular Representations)**.** *Suppose that V is a finite-dimensional vector space of dimension m and $T : V \to V$ is linear. Then there is a chain $\{0\} = V_0 \subseteq V_1 \subset \cdots \subset V_m = V$ such that $\dim(V_i) = i$ and such that each V_i is T-invariant.*

Proof. We prove the theorem by induction on m, the case $m = 0$ being trivial. Assume now that $m \geq 1$ and the theorem is true for vector spaces of dimension $m - 1$. Suppose $\dim(V) = m$ and $T : V \to V$ is linear. Since we are working over $\mathbb{C}$, we can find an eigenvector u of T. Consider a basis $\{u, w_1, \ldots, w_{m-1}\}$ of V. Set $U = \mathrm{sp}(u)$, $W = \mathrm{sp}(w_1, \ldots, w_{m-1})$ and $P : V \to W$ the projection onto W. Define $S : W \to W$ by $S(w) = P(T(w))$. By the induction hypothesis, we have a basis $\{z_1, \ldots, z_{m-1}\}$ for W such that, for each $i \in \{1, \ldots, m-1\}$, $S(\mathrm{sp}(z_1, \ldots, z_i)) \subseteq \mathrm{sp}(z_1, \ldots, z_i)$. We leave it to the reader to check that $T(\mathrm{sp}(u, z_1, \ldots, z_i)) \subseteq \mathrm{sp}(u, z_1, \ldots, z_i)$ for each $i \in \{1, \ldots, m-1\}$. Set $V_1 := U$ and, for $i = 1, \ldots, m-1$, $V_{i+1} := \mathrm{sp}(u, z_1, \ldots, z_i)$. □

We now apply the (transfer of the) previous theorem to H_ν and $\tilde{T}$, yielding a chain

$$\{0\} = E_0 \subset E_1 \subset \cdots \subset E_\nu = H_\nu,$$

where each $E_i \in \mathcal{E}^*$, $\dim(E_i) = i$, and each E_i is $\tilde{T}$-invariant. For appropriately chosen $\{e_1, e_2, \ldots\}$ and $\nu \in \mathbb{N}^* \setminus \mathbb{N}$, we will have that each $\mathrm{st}(E_i)$ is T-invariant. Since each $\mathrm{st}(E_i)$ is closed, it will remain to find some $i \in \{1, \ldots, \nu\}$ such that $\mathrm{st}(E_i)$ is nontrivial.

Let's get started: fix $v \in H$, $\|v\| = 1$. Let $A = \{v, Tv, T^2v, \ldots\}$. Let's eliminate some trivial cases early on. First, suppose that A is linearly dependent. Then there are $\alpha_0, \alpha_1, \ldots, \alpha_{k-1} \in \mathbb{C}$, not all 0, such that $T^kv = \alpha_0 v + \alpha_1 Tv + \cdots + \alpha_{k-1}T^{k-1}v$. Set $Y := \mathrm{sp}(v, Tv, \ldots, T^{k-1}v)$. Then Y is finite-dimensional (hence closed), nontrivial since $v \in Y$ and H is infinite-dimensional, and T-invariant.

We may thus assume that A is linearly independent. Let F be the closed linear span of A. Suppose that $F \neq H$. We claim that F is T-invariant, whence F is a nontrivial T-invariant subspace of H. First, suppose that $w \in \mathrm{sp}(A)$, so $w = \alpha_0 v + \alpha_1 Tv + \cdots + \alpha_k T^k v$ for some $\alpha_0, \ldots, \alpha_k \in \mathbb{C}$. But then $Tw = \alpha_0 Tv + \alpha_1 T^2 v + \cdots + \alpha_k T^{k+1}v \in \mathrm{sp}(A)$. For the more general situation, $w = \lim_{n\to\infty} w_n$ with each $w_n \in \mathrm{sp}(A)$. But then, since T is continuous, $Tw = \lim_{n\to\infty} Tw_n \in F$ since $Tw_n \in \mathrm{sp}(A)$ for each n.

We may thus suppose $F = H$. In this case, we apply the Gram-Schmidt process to A to get an orthonormal basis $\{e_1, e_2, \ldots\}$ of H with the property that $\mathrm{sp}(x, Tx, \ldots, T^{n-1}x) = \mathrm{sp}(e_1, e_2, \ldots, e_n)$ for each $n \geq 1$.

Let's observe something interesting about the "matrix representation" of T with respect to the basis $\{e_1, e_2, \ldots\}$. For $k \in \mathbb{N}^{>0}$, we have $e_k \in \mathrm{sp}(x, Tx, \ldots, T^{k-1}x)$, so $T(e_k) \in \mathrm{sp}(Tx, T^2x, \ldots, T^kx) \subseteq \mathrm{sp}(e_1, \ldots, e_{k+1})$. Thus, if we write $T(e_k) = \sum_{j=1}^\infty a_{jk}e_k$, we see that $a_{jk} = 0$ for $j > k+1$. We then say that the matrix $[a_{jk}]$ for T is *almost superdiagonal.*

It will prove desirable later on to know that T and $\tilde{T}$ are "close." More precisely, we will want to know that, for $y \in H_\nu \cap H_{\mathrm{fin}}$, $T(y) \approx \tilde{T}(y)$. We can achieve this goal by choosing ν appropriately, as we proceed to show now.

First, we would like to know a predictable form for the matrix representation of T^n with respect to our basis $\{e_1, e_2, \ldots\}$. Usually, this is not possible, but thankfully, $[a_{jk}]$ is almost superdiagonal. Recursively define $a_{jk}^{(n)}$ as follows: Set $a_{jk}^{(1)} := a_{jk}$ and $a_{jk}^{(n+1)} := \sum_{i=1}^{\infty} a_{ji}^{(n)} a_{ik} = \sum_{i=1}^{k+1} a_{ji}^{(n)} a_{ik}$.

Exercise 14.12. For $n, j, k > 0$, we have:

(1) $a_{j+k,j}^{(n)} = 0$ if $n < k$, and
(2) $a_{j+k,j}^{(k)} = \prod_{i=0}^{k-1} a_{j+i+1,j+i}$.

For the rest of this section, fix a polynomial $p(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_k z^k$ with coefficients from $\mathbb{C}$ such that $\alpha_k \neq 0$ and $p(T)$ is compact. We can now choose the "appropriate" $\nu \in \mathbb{N}^* \setminus \mathbb{N}$.

Lemma 14.13. *There is $\nu \in \mathbb{N}^* \setminus \mathbb{N}$ such that $a_{\nu+1,\nu} \approx 0$.*

Proof. Let $[b_{jk}]$ be the matrix for $p(T)$. Then by the previous two exercises,

$$\begin{aligned} b_{j+k,j} &= 0 + \alpha_1 a_{j+k,j}^{(1)} + \alpha_2 a_{j+k,j}^{(2)} + \cdots + \alpha_k a_{j+k,j}^{(k)} \\ &= \alpha_k a_{j+k,j}^{(k)} \\ &= \alpha_k \prod_{i=0}^{k-1} a_{j+i+1,j+i}. \end{aligned}$$

By Lemma 11.25, $b_{N+k,N} \approx 0$ for all $N > \mathbb{N}$. Thus, there is $i \in \{0, \ldots, k-1\}$ such that $a_{N+i+1,N+i} \approx 0$. Set $\nu = N + i$ for this i. $\square$

For the remainder of this section, we fix ν as in the previous lemma. Let us show how this ν has the promised property.

Lemma 14.14. *If $y \in H_\nu \cap H_{\mathrm{fin}}$, then $T(y) \approx \tilde{T}(y)$.*

Proof. Write $y = \sum_{k=1}^{\nu} \alpha_k e_k$ with each $\alpha_k \in \mathbb{C}^*$. By the Pythagorean Theorem, $\|y\|^2 = \sum_{k=1}^{\nu} |\alpha_k|^2 \geq |\alpha_\nu|^2$, whence $\alpha_\nu \in \mathbb{C}_{\mathrm{fin}}$ since $y \in H_{\mathrm{fin}}$. Using that $[a_{jk}]$ is almost superdiagonal, we have

$$\begin{aligned} T(y) = \sum_{k=1}^{\nu} \alpha_k T(e_k) &= \sum_{k=1}^{\nu} \alpha_k \sum_{j=1}^{k+1} a_{jk} e_j \\ &= \sum_{k=1}^{\nu} \alpha_k \sum_{j=1}^{\nu+1} a_{jk} e_j \\ &= \sum_{j=1}^{\nu+1} (\sum_{k=1}^{\nu} \alpha_k a_{jk}) e_j. \end{aligned}$$

Thus, $\tilde{T}(y) = P_\nu T(y) = \sum_{j=1}^{\nu}(\sum_{k=1}^{\nu} \alpha_k a_{jk})e_j$, whence

$$\begin{aligned} T(y) &= P_\nu T(y) + \sum_{k=1}^{\nu} \alpha_k a_{\nu+1,k} e_{\nu+1} \\ &= P_\nu T(y) + \alpha_\nu a_{\nu+1,\nu} e_{\nu+1} \\ &\approx P_\nu T(y), \end{aligned}$$

since $a_{\nu+1,\nu} \approx 0$ and $\alpha_\nu \in \mathbb{C}_{\text{fin}}$. □

Corollary 14.15. *For any $n \in \mathbb{N}^{>0}$ and $y \in H_\nu \cap H_{\text{fin}}$, we have $T^n(y) \approx \tilde{T}^n(y)$. Consequently, for any polynomial $q(z)$ and any $y \in H_\nu \cap H_{\text{fin}}$, $q(T)y \approx q(\tilde{T})(y)$.*

Proof. The proof is by induction on n, the case $n = 1$ being the content of the previous lemma. For the induction step, assume that $n > 1$ and that the corollary holds for $n - 1$. Since $\tilde{T}$ is internally bounded and $y \in H_{\text{fin}}$, we have that $\tilde{T}^{n-1}(y) \in H_{\text{fin}}$. Consequently,

$$T^n(y) = T(T^{n-1}(y)) \approx T(\tilde{T}^{n-1}(y)) \approx \tilde{T}(\tilde{T}^{n-1}(y)) = \tilde{T}^n(y),$$

where the first $\approx$ holds by the continuity of T and the inductive assumption that $T^{n-1}(y) \approx \tilde{T}^{n-1}(y)$, while the second $\approx$ holds by the base case of the induction applied to the vector $\tilde{T}^{n-1}(y)$. □

Lemma 14.16. *For $i = 0, 1, \ldots, \nu$, we have $T(\text{st}(E_i)) \subseteq \text{st}(E_i)$.*

Proof. Fix $i \in \{0, 1, \ldots, \nu\}$ and set $E := E_i$. Fix $x \in \text{st}(E)$; we need $T(x) \in \text{st}(E)$. By Exercise 11.47, we have $P_E(x) \approx x$. Since T is continuous, $TP_E(x) \approx Tx$. Consequently, $TP_E(x) \in H_{\text{ns}}$. Thus, by Lemma 13.6, $P_\nu(TP_E(x)) \approx TP_E(x)$. Consequently, we have $T_\nu P_E(x) = P_\nu T P_\nu P_E(x) = P_\nu T P_E(x) \approx TP_E(x) \approx Tx$. Since $T_\nu(E) = \tilde{T}(E) \subseteq E$, it follows that $T(x) \in \text{st}(E)$. □

It thus remains to find some i such that $\text{st}(E_i)$ is nontrivial. To achieve this goal, we define the internal sequence $r_0, r_1, \ldots, r_\nu$ of numbers from $\mathbb{R}^*$ by $r_j := \|p(T_\nu)(v) - p(T_\nu)P_{E_j}(v)\|$. (Recall that we have fixed a vector $v \in H$ of norm 1.) We first show that $r_0 \not\approx 0$ while $r_\nu \approx 0$. To do this, we first prove:

Lemma 14.17. *$p(T_\nu)$ is S-continuous.*

Proof. Observe that

$$\|p(T_\nu)\| \leq \sum_{i=0}^{k} |\alpha_i| \|T_\nu\|^i \leq \sum_{i=0}^{k} |\alpha_i| \|T\|^i,$$

whence $\|p(T_\nu)\| \in \mathbb{R}_{\text{fin}}$. Now suppose that $x, y \in H^*$ and $x \approx y$. Then $\|p(T_\nu)x - p(T_\nu)y\| \leq \|p(T_\nu)\| \|x - y\| \approx 0$. □

For simplicity, we write P_j instead of P_{E_j}.

Corollary 14.18. *$r_0 \not\approx 0$ and $r_\nu \approx 0$.*

Proof. By Lemma 13.6, $P_\nu(v) \approx v$. Since $p(T)$ is continuous, $p(T)(P_\nu(v)) \approx p(T)(v)$. However, since $P_\nu(v) \in H_\nu$, we have

$$p(T)(P_\nu(v)) \approx p(\tilde{T})(P_\nu(v)) = p(T_\nu)(P_\nu(v)) \approx p(T_\nu)(v).$$

Thus, $p(T_\nu)(v) \approx p(T)(v) \neq 0$ since $\{v, Tv, T^2v, \ldots\}$ is linearly independent. Since $E_0 = \{0\}$, $r_0 = \|p(T_\nu(v))\| \not\approx 0$. On the other hand,

$$r_\nu = \|p(T_\nu)(v) - p(T_\nu)P_\nu(v)\| \approx 0$$

since $P_\nu(v) \approx v$ and $p(T_\nu)$ is S-continuous.

□

Observe that in the previous two results, we did not use the fact that $p(T)$ is compact, that is, the previous two results are valid for any polynomial $q(z)$.

Choose $r \in \mathbb{R}^{>0}$ such that $r < r_0$. Set η to be the least $k \in \{0, 1, \ldots, \nu\}$ such that $r_k < \frac{1}{2}r$; since $r_\nu \approx 0$, $r_\nu < \frac{1}{2}r$, so $0 < \eta \leq \nu$. We will now show that either $\mathrm{st}(E_{\eta-1})$ is nontrivial or $\mathrm{st}(E_\eta)$ is nontrivial.

Lemma 14.19. *$v \notin \mathrm{st}(E_{\eta-1})$ (whence $\mathrm{st}(E_{\eta-1}) \neq H$).*

Proof. Suppose that $v \in \mathrm{st}(E_{\eta-1})$. Then by Exercise 11.47, $v \approx P_{\eta-1}v$. But then, by Lemma 14.17, $p(T_\nu)v \approx p(T_\nu)P_{\eta-1}(v)$, whence $r_{\eta-1} \approx 0$. But $r_{\eta-1} \geq \frac{1}{2}r$, a contradiction. □

Lemma 14.20. $\mathrm{st}(E_\eta) \neq \{0\}$.

Proof. Set $y := p(T_\nu)P_\eta(v)$. Since E_η is $\tilde{T}$-invariant, it is also $p(\tilde{T})$-invariant, whence $y \in E_\eta$. By Lemma 14.15, $y = p(T_\nu)P_\eta(v) \approx p(T)P_\eta(v)$. Since $p(T)$ is compact and $P_\eta(v) \in H_{\mathrm{fin}}$, we have that $p(T)P_\eta(v) \in H_{\mathrm{ns}}$, whence $y \in H_{\mathrm{ns}}$. Let $z := \mathrm{st}(y) \in H$, whence $z \in \mathrm{st}(E_\eta)$. If $z = 0$, then $y \approx 0$, whence $r_\eta \approx \|p(T_\nu)(v)\| = r_0 > r$. This contradicts the fact that $r_\eta < \frac{1}{2}r$. □

Thus, the only way for both $\mathrm{st}(E_{\eta-1})$ and $\mathrm{st}(E_\eta)$ to both be trivial is for $\mathrm{st}(E_{\eta-1}) = \{0\}$ and $\mathrm{st}(E_\eta) = H$. We now show that this cannot happen for "dimension reasons."

Lemma 14.21. *Suppose that $E, F \in \mathcal{E}^*$, $E \subseteq F$, and $\dim(F) = \dim(E)+1$. Then for any $x, y \in \mathrm{st}(F)$, there is $\lambda \in \mathbb{C}$ and $z \in \mathrm{st}(E)$ such that $x = \lambda y + z$ or $y = \lambda x + z$.*

Proof. Fix $x', y' \in F$ such that $x \approx x'$ and $y \approx y'$. By the transfer principle, there is $\lambda \in \mathbb{C}^*$ and $z \in E$ such that either $y' = \lambda x' + z$ or $x' = \lambda y' + z$. Without loss of generality, we assume $y' = \lambda x' + z$. If $\lambda \in \mathbb{C}_{\mathrm{fin}}$, then since $z = y' - \lambda x' \in E$, we have $z \in H_{\mathrm{ns}}$ and $\mathrm{st}(z) \in \mathrm{st}(E)$. If $\lambda \notin \mathbb{C}_{\mathrm{fin}}$, then $\frac{1}{\lambda}y' = x' + \frac{1}{\lambda}z$, reverting us back to the previous case. □

Suppose, towards a contradiction, that $\mathrm{st}(E_{\eta-1}) = \{0\}$ and $\mathrm{st}(E_\eta) = H$. By the previous lemma, for any $x, y \in H$, there is $\lambda \in \mathbb{C}$ such that $x = \lambda y$ or $y = \lambda x$, that is, $\dim(H) = 1$, a *serious* contradiction. □

15. Measure Theory

In this section, we will describe the Loeb measure construction. There is a lot to say here, but we will cover only enough to discuss the application to combinatorics in the final section.

15.1. **General measure theory.** Fix a set X. A nonempty set $\Omega \subseteq \mathcal{P}(X)$ is an *algebra* if it is closed under unions, intersections, and complements, that is, if $A, B \in \Omega$, then $A \cup B$, $A \cap B$, and $X \setminus A$ all belong to Ω. If Ω is an algebra of subsets of X, then $\emptyset, X \in \Omega$. An algebra Ω on X is said to be a *σ-algebra* if it is also closed under countable unions, that is, if $A_1, A_2, \ldots$ all belong to Ω, then so does $\bigcup_{n=1}^{\infty} A_n$. A σ-algebra is then automatically closed under countable intersections.

Exercise 15.1. Suppose that X is a set and $\mathcal{O} \subseteq \mathcal{P}(X)$ is an arbitrary collection of subsets of X. Prove that there is a smallest σ-algebra Ω containing $\mathcal{O}$. We call this σ-algebra the *σ-algebra generated by $\mathcal{O}$* and denote it by $\sigma(\mathcal{O})$.

Remark. When trying to prove that every element of $\sigma(\mathcal{O})$ has a certain property, one just needs to show that the set of elements having that property contains $\mathcal{O}$ and is a σ-algebra.

Suppose that Ω is an algebra on X. A *pre-measure* on Ω is a function $\mu : \Omega \to [0, +\infty]$ satisfying the following two axioms:

- $\mu(\emptyset) = 0$;
- (Countable Additivity) If $A_1, A_2, \ldots,$ all belong to Ω, are pairwise disjoint, *and* $\bigcup_{n=1}^{\infty} A_n$ belongs to Ω, then $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

If Ω is a σ-algebra, then a pre-measure is called a *measure*. If μ is a measure on X and $\mu(X) = 1$, then we call μ a *probability measure on X*.

Example 15.2. Fix $n \in \mathbb{N}$ and suppose that $X = \{1, 2, \ldots, n\}$. Let $\Omega := \mathcal{P}(X)$. Then Ω is an algebra of subsets of X that is actually a σ-algebra for trivial reasons. Define the function $\mu : \Omega \to [0, 1]$ by $\mu(A) = \frac{|A|}{n}$. Then μ is a probability measure on Ω, called the *normalized counting measure*.

Exercise 15.3. Suppose that $\mu : \Omega \to [0, +\infty]$ is a pre-measure. Prove that $\mu(A) \leq \mu(B)$ for all $A, B \in \Omega$ with $A \subseteq B$.

For subsets A, B of X, we define the *symmetric difference of A and B* to be $A \triangle B := (A \setminus B) \cup (B \setminus A)$.

Exercise 15.4. Suppose that Ω is an algebra and $\mu : \sigma(\Omega) \to [0, \infty]$ is a measure. Prove that, for any $A \in \sigma(\Omega)$ with $\mu(A) < \infty$ and any $\epsilon \in \mathbb{R}^{>0}$, there is $B \in \Omega$ such that $\mu(A \triangle B) < \epsilon$.

For our purposes, it will be of vital importance to know that a pre-measure μ on an algebra Ω can be extended to a measure on a σ-algebra Ω' extending Ω. We briefly outline how this is done; the interested reader can consult any good book on measure theory for all the glorious details.

Fix an algebra Ω of subsets of X and a pre-measure μ on Ω. For arbitrary $A \subseteq X$, we define the *outer measure of A* to be

$$\mu^+(A) := \inf\{\sum_{n\in\mathbb{N}} \mu(B_n) \mid A \subseteq \bigcup_{n\in\mathbb{N}} B_n, \text{ each } B_n \in \Omega\}.$$

Note that $\mu^+(A) = \mu(A)$ for all $A \in \Omega$.

Now although μ^+ is defined on all of $\mathcal{P}(X)$ (which is certainly a σ-algebra), it need not be a measure. However, there is a canonical σ-sub-algebra Ω_m of $\mathcal{P}(X)$, the so-called *μ^+-measurable subsets of X*, on which μ^+ is a measure. (The precise definition of μ^+-measurable is not important here; again, any standard measure theory text will have the details.) Let us collect the relevant facts here:

Fact 15.5. *Let X be a set, Ω an algebra of subsets of X, and $\mu : \Omega \to [0,\infty]$ a pre-measure on Ω with associated outer measure μ^+ and σ-algebra of μ^+-measurable sets Ω_m. Further suppose that μ is* σ-finite, *meaning that we can write $X = \bigcup_{n\in\mathbb{N}} X_n$ with each $X_n \in \Omega$ and $\mu(X_n) < \infty$.*

1. *$\sigma(\Omega) \subseteq \Omega_m$ and $\mu^+|\Omega = \mu$.*
2. *(Uniqueness) If Ω' is another σ-algebra on X extending Ω and $\mu' : \Omega' \to [0,\infty]$ is a measure on Ω' extending μ, then μ^+ and μ' agree on $\Omega_m \cap \Omega'$ (and, in particular, on $\sigma(\Omega)$).*
3. *(Completeness) If $A \subseteq B \subseteq X$ are such that $B \in \Omega_m$ and $\mu^+(B) = 0$, then $A \in \Omega_m$ and $\mu^+(A) = 0$.*
4. *(Approximation Results)*
   (a) *If $A \in \Omega_m$, then there is $B \in \sigma(\Omega)$ containing A such that $\mu^+(B \setminus A) = 0$. (So Ω_m is the* completion *of $\sigma(\Omega)$.)*
   (b) *If $A \in \Omega_m$ is such that $\mu^+(A) < \infty$, then for every $\epsilon \in \mathbb{R}^{>0}$, there is $B \in \Omega$ such that $\mu(A \triangle B) < \epsilon$.*
   (c) *Suppose that $A \subseteq X$ is such that, for every $\epsilon \in \mathbb{R}^{>0}$, there is $B \in \Omega$ such that $\mu(A \triangle B) < \epsilon$. Then $A \in \Omega_m$.*

Example 15.6 (Lebesgue measure)**.** Suppose that $X = \mathbb{R}$ and Ω is the collection of *elementary sets*, namely the finite unions of intervals. Define $\mu : \Omega \to [0,\infty]$ by declaring $\mu(I) = \text{length}(I)$ and $\mu(I_1 \cup \cdots \cup I_n) = \sum_{i=1}^n \mu(I_j)$ whenever $I_1, \ldots, I_n$ are pairwise disjoint. The above outer-measure procedure yields the σ-algebra Ω_m, which is known as the σ-algebra of *Lebesgue measurable subsets of* $\mathbb{R}$ and usually denoted by $\mathfrak{M}$. The measure μ^+ is often denoted by λ and is referred to as *Lebesgue measure*. The σ-algebra $\sigma(\Omega)$ in this case is known as the σ-algebra of *Borel subsets of* $\mathbb{R}$. It can also be seen to be the σ-algebra generated by the open intervals.

15.2. **Loeb measure.** How do we obtain pre-measures in the nonstandard context? Well, we obtain them by looking at normalized counting measures on hyperfinite sets. Suppose that X is a hyperfinite set. We set Ω to be the set of *internal* subsets of X. Then Ω is an algebra of subsets of X that is not (in general) a σ-algebra. For example, if $X = \{0, 1, \ldots, N\}$ for some

$N > \mathbb{N}$, then for each $n \in \mathbb{N}$, $A_n := \{n\}$ belongs to Ω, but $\bigcup_n A_n = \mathbb{N}$ does not belong to Ω as $\mathbb{N}$ is not internal.

If $A \in \Omega$, then A is also hyperfinite. We thus define a function $\mu : \Omega \to [0,1]$ by $\mu(A) := \mathrm{st}(\frac{|A|}{|X|})$. We claim that μ is a pre-measure. It is easily seen to be *finitely additive*, that is, $\mu(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^n \mu(A_i)$ whenever $A_1, \ldots, A_n \in \Omega$ are disjoint. But how do we verify countable additivity?

Exercise 15.7. If $A_1, A_2, \ldots$ all belong to Ω and $\bigcup_{n=1}^\infty A_n$ also belongs to Ω, then there is $k \in \mathbb{N}$ such that $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^k A_n$.

Thus, by the exercise, countable additivity is a trivial consequence of finite additivity in this context. We may thus apply the outer measure procedure in the previous section to obtain a probability measure $\mu^+ : \Omega_m \to [0,1]$ extending μ. μ^+ is called the *Loeb measure on* X and is often denoted μ_L. The elements of Ω_m are referred to as the *Loeb measurable subsets of* X and this set is usually denoted by Ω_L.

There are many interesting things to say about Loeb measure. It is crucial for applications of nonstandard analysis to many different areas of mathematics. Let us just mention its connection to Lebesgue measure. We will prove the following theorem in the exercises.

Theorem 15.8. *Suppose that $N > \mathbb{N}$ and $X = \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N} = 1\}$, a hyperfinite set. Consider* $\mathrm{st} : X \to [0,1]$. *Define a σ-algebra Ω on $[0,1]$ by $A \in \Omega$ if and only if* $\mathrm{st}^{-1}(A) \in \Omega_L$. *For $A \in \Omega$, define $\nu(A) := \mu_L(\mathrm{st}^{-1}(A))$. Then $\Omega = \mathfrak{M}$ and $\nu = \lambda$. (In other words, Lebesgue measure can be constructed nonstandardly from a hyperfinite normalized counting measure.)*

15.3. **Product measure.** Suppose that (X, Ω_X, μ_X) and (Y, Ω_Y, μ_Y) are two probability measure spaces. We can then form their *product* as follows: first, set Ω to be the set of finite unions of rectangles of the form $A \times B$, where $A \in \Omega_X$ and $B \in \Omega_Y$. It is an exercise to show that Ω is an algebra of subsets of $X \times Y$ and that every element of Ω can be written as a finite union of *disjoint* such rectangles. We can then define a pre-measure μ on Ω by $\mu(\bigcup_{i=1}^n (A_i \times B_i)) := \sum_{i=1}^n (\mu_X(A_i) \cdot \mu_Y(B_i))$. Applying the outer measure procedure, we get a measure $\mu_X \otimes \mu_Y : \Omega_m \to [0,1]$ extending μ. We denote Ω_m by $\Omega_X \otimes \Omega_Y$.

The following situation will come up in the final section: suppose that X and Y are hyperfinite sets and we construct the Loeb measure spaces $(X, \Omega_{X,L}, \mu_{X,L})$ and $(Y, \Omega_{Y,L}, \mu_{Y,L})$. We are thus entitled to consider the product measure space $(X \times Y, \Omega_{X,L} \otimes \Omega_{Y,L}, \mu_{X,L} \otimes \mu_{Y,L})$. However, $X \times Y$ is itself a hyperfinite set, whence we can consider its Loeb measure space $(X \times Y, \Omega_{X \times Y,L}, \mu_{X \times Y,L})$. There is a connection:

Exercise 15.9. Show that $\Omega_{X,L} \otimes \Omega_{Y,L}$ is a sub-σ-algebra of $\Omega_{X \times Y,L}$ and that $\mu_{X \times Y,L} | (\Omega_{X,L} \otimes \Omega_{Y,L}) = \mu_{X,L} \otimes \mu_{Y,L}$.

15.4. **Integration.** Once one has a measure space (X, Ω, μ), one can integrate as follows. First, for a function $f : X \to \mathbb{R}$, we say that f is *measurable* if $f^{-1}(U) \in \Omega$ for any open $U \subseteq \mathbb{R}$.

Next, define a *simple* function $g : X \to \mathbb{R}$ to be a measurable function with finite range. Given a set $A \subseteq X$, we define $1_A : X \to \mathbb{R}$ by $1_A(x) = 1$ if $x \in A$ and 0 otherwise. If $A \in \Omega$, then 1_A is measurable. Any simple function g can be written as $g = \sum_{i=1}^{n} r_i 1_{A_i}$ with $r_i \in \mathbb{R}$ and $A_i \in \Omega$. For such a simple function g, we define the *integral of* g to be $\int g d\mu := \sum_{i=1}^{n} r_i \mu(A_i)$. For an arbitrary *positive* measurable function $f : X \to \mathbb{R}$, we define the *integral of* f to be $\int f d\mu := \sup\{\int g d\mu \mid g \leq f,\ g \text{ a simple function}\}$. f is said to be *integrable* if $\int f d\mu < \infty$. For an arbitrary function f, we set $f^+ := \max(f, 0)$ and $f^- := \max(-f, 0)$. We say that f is integrable if both f^+ and f^- are integrable, in which case we define $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$.

If $A \in \Omega$, then we write $\int_A f d\mu$ to indicate $\int f \cdot 1_A d\mu$ and call this the integral of f on A.

15.5. **Conditional expectation.** Let (X, Ω, μ) be a measure space. We define an inner-product space $L^2(X, \Omega, \mu)$ as follows: the set of vectors is the set of measurable functions $f : X \to \mathbb{R}$ such that $\int |f|^2 d\mu < \infty$. We add vectors and multiply by scalars in the usual way. We define the inner product to be $\langle f, g \rangle := \int f g d\mu$. Of course, one has to check that this integral is finite and this satisfies the requirements of an inner product, but this follows from the appropriate properties of the integral. (O.K. We are telling a bit of a lie. Technically speaking, we should be considering equivalence classes of measurable functions with respect to the equivalence relation $f \equiv g$ if $\{x \in X \mid f(x) \neq g(x)\}$ is contained in a set in Ω of measure 0. We will ignore this important subtlety.)

Theorem 15.10 (Riesz-Fisher Theorem)**.** *$L^2(X, \Omega, \mu)$ is a Hilbert space.*

The proof of the Riesz-Fisher Theorem requires some basic convergence theorems for integrals, but we will not get into this now.

Now suppose that Ω' is a sub-σ-algebra of Ω. We then get a subspace $L^2(X, \Omega', \mu)$ of $L^2(X, \Omega, \mu)$ consisting of the Ω'-measurable functions.

Exercise 15.11. $L^2(X, \Omega', \mu)$ is a closed subspace of $L^2(X, \Omega, \mu)$.

We thus have the orthogonal projection $P : L^2(X, \Omega, \mu) \to L^2(X, \Omega', \mu)$. Consider $A \in \Omega'$ with $\mu(A) < \infty$, so $1_A \in L^2(X, \Omega', \mu)$. Then for any $g \in L^2(X, \Omega, \mu)$, we have $\langle g, 1_A \rangle = \langle P(g), 1_A \rangle$, or, in integral notation:

$$\int_A g d\mu = \int_A P(g) d\mu.$$

In other words, $P(g)$ is an Ω'-measurable function which has the same integral as g over subsets of Ω'. For probability reasons, $P(g)$ is called the *conditional expectation of g with respect to* Ω' and is often denoted $\mathbb{E}[g|\Omega']$.

15.6. **Problems.** We are aiming to prove the following:

Theorem: Suppose that $N > \mathbb{N}$ and $X = \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N} = 1\}$, a hyperfinite set. Consider $\mathrm{st} : X \to [0,1]$. Define a σ-algebra Ω on $[0,1]$ by $A \in \Omega$ if and only if $\mathrm{st}^{-1}(A) \in \Omega_L$. For $A \in \Omega$, define $\nu(A) := \mu_L(\mathrm{st}^{-1}(A))$. Then $\Omega = \mathfrak{M}$ and $\nu = \lambda$.

We will do this in steps.

Problem 15.1. Prove that Ω is a σ-algebra and ν is a measure on Ω.

Problem 15.2. Fix $a, b \in [0,1]$ with $a < b$.

(1) Prove that $X \cap (a,b)^* \in \Omega_L$ and $\mu_L(X \cap (a,b)^*) = b - a$.
(2) Prove that $\mathrm{st}^{-1}((a,b)) = \bigcup_{n\in\mathbb{N}}(X \cap (a + \frac{1}{n}, b - \frac{1}{n})^*)$.
(3) Prove that $(a,b) \in \Omega$ and $\nu((a,b)) = b - a$.

Let $\mathcal{B}$ denote the σ-algebra of Borel subsets of $[0,1]$. We now use the fact that λ is the only measure on $\mathcal{B}$ satisfying $\lambda(a,b) = b - a$ to conclude that $\mathcal{B} \subseteq \Omega$ and $\nu|\mathcal{B} = \lambda|\mathcal{B}$.

Problem 15.3. Conclude that $\mathfrak{M} \subseteq \Omega$ and $\nu|\mathfrak{M} = \lambda|\mathfrak{M}$. (Hint: Use Fact 15.5.)

It remains to show that $\Omega \subseteq \mathfrak{M}$. We will need the following fact, whose proof we will return to after we finish the Lebesgue measure business:

Fact 15.12. *If $B \in \Omega_X$, then, for every $\epsilon \in \mathbb{R}^{>0}$, there are internal subsets C, D of X such that $C \subseteq B \subseteq D$ and $\mu(D \setminus C) < \epsilon$.*

Now suppose that $B \in \Omega$. Fix $\epsilon \in \mathbb{R}^{>0}$. By the previous fact, there are internal $C, D \subseteq X$ such that $C \subseteq \mathrm{st}^{-1}(B) \subseteq D$ and $\mu(D \setminus C) < \epsilon$. Set $C' := \mathrm{st}(C)$ and $D' := [0,1] \setminus \mathrm{st}(X \setminus D)$. Notice that C' is closed and D' is open (why?), whence $C', D' \in \mathcal{B} \subseteq \Omega$.

Problem 15.4. Prove that $C \subseteq \mathrm{st}^{-1}(C')$ and $\mathrm{st}^{-1}(D') \subseteq D$. Conclude that $B \in \mathfrak{M}$.

We now explain how to prove Fact 15.12. Recall that the outer measure procedure yields the following formula for μ_L:

$$\mu_L(B) = \inf\{\sum_{n\in\mathbb{N}} \mu(A_n) \mid \text{ each } A_n \text{ is internal and } B \subseteq \bigcup_{n\in\mathbb{N}} A_n\}.$$

Problem 15.5. In this problem, we aim to prove, if $B \in \Omega_L$, then

$$\mu_L(B) = \inf\{\mu_L(A) \mid A \text{ is internal and } B \subseteq A\}.$$

The inequality $\leq$ is clear. Fix $\epsilon \in \mathbb{R}^{>0}$; we need to find internal A such that $B \subseteq A$ and $\mu_L(A) \leq \mu_L(B) + \epsilon$. Fix internal sets $(A_n \mid n \in \mathbb{N})$ such that $B \subseteq \bigcup_{n\in\mathbb{N}} A_n$ and $\mu_L(\bigcup_{n\in\mathbb{N}} A_n) < \mu_L(B) + \epsilon$. Without loss of generality, we may assume that $A_n \subseteq A_{n+1}$ for all n. By countable comprehension, we extend this sequence to an internal sequence $(A_n \mid n \in \mathbb{N}^*)$.

(1) Prove that, for each $k \in \mathbb{N}$, we have
$$(\forall n \in \mathbb{N}^*)(n \leq k \to (A_n \subseteq A_k \text{ and } \mu(A_n) < \mu_L(B) + \epsilon)).$$
(2) Prove that there is $K > \mathbb{N}$ such that $\mu_L(A_K) \leq \mu_L(B) + \epsilon$, finishing the proof.

Problem 15.6. Prove a dual version of the previous problem, namely, if $B \in \Omega_L$, then
$$\mu_L(B) = \sup\{\mu_L(A) \mid A \text{ is internal and } A \subseteq B\}.$$

Problem 15.7. Use the previous two problems to prove Fact 15.12.

16. Szemerédi Regularity Lemma

A nice recent application of nonstandard methods is a slick proof of the celebrated *Szemerédi Regularity Lemma* for graphs. To state this result, we need to introduce some terminology. Suppose that (V, E) is a finite graph. For two disjoint, nonempty subsets X, Y of V, we define the *density of arrows between X and Y* to be the quantity
$$\delta(X, Y) := \frac{|E \cap (X \times Y)|}{|X||Y|}.$$
For example, if every element of X is connected to every element of Y by an edge, then $\delta(X, Y) = 1$. Fix $\epsilon \in \mathbb{R}^{>0}$. We say that X and Y as above are *ϵ-pseudorandom* if whenever $A \subseteq X$ and $B \subseteq Y$ are such that $|A| \geq \epsilon|X|$ and $|B| \geq \epsilon|Y|$, then $|\delta(A, B) - \delta(X, Y)| < \epsilon$. In other words, as long as A and B contain at least an ϵ proportion of the elements of X and Y, then $\delta(A, B)$ is essentially the same as $\delta(X, Y)$, so the edges between X and Y are distributed in a sort of random fashion.

If $X = \{x\}$ and $Y = \{y\}$ are singletons, then clearly X and Y are ϵ-pseudorandom for any ϵ. Thus, any finite graph can trivially be partitioned into a finite number of ϵ-pseudorandom pairs by partitioning the graph into singletons. Szemerédi's Regularity Lemma essentially says that one can do much better in the sense that there is a constant $C(\epsilon)$ such that any finite graph has an "ϵ-pseudorandom partition" into at most $C(\epsilon)$ pieces. Unfortunately, the previous sentence isn't 100% accurate; there's a bit of error that we need to account for.

Suppose that $V_1, \ldots, V_m$ is a partition of V into m pieces. Set
$$R := \{(i, j) \mid 1 \leq i, j \leq m, \quad V_i \text{ and } V_j \text{ are } \epsilon\text{-pseudorandom}\}.$$
We say that the partition is *ϵ-regular* if $\sum_{(i,j)\in R} \frac{|V_i||V_j|}{|V|^2} > (1 - \epsilon)$. This says that, in some sense, almost all of the pairs of points are in ϵ-pseudorandom pairs. We can now state

Theorem 16.1 (Szemerédi's Regularity Lemma)**.** *For any $\epsilon \in \mathbb{R}^{>0}$, there is a constant $C(\epsilon)$ such that any graph (V, E) admits an ϵ-regular partition into $m \leq C(\epsilon)$ pieces.*

Exercise 16.2. Prove that Szemerédi's Regularity Lemma is equivalent to the following statement: for any $\epsilon \in \mathbb{R}^{>0}$ and any hyperfinite graph (V, E), there is a *finite* partition $V_1, \ldots, V_m$ of V into internal sets and a subset $R \subseteq \{1, \ldots, m\}^2$ such that:

- for $(i, j) \in R$, V_i and V_j are *internally* ϵ-pseudorandom: for all *internal* $A \subseteq V_i$ and $B \subseteq V_j$ with $|A| \geq \epsilon |V_i|$ and $|B| \geq \epsilon |V_j|$, we have $|\delta(A, B) - \delta(V_i, V_j)| < \epsilon$; and
- $\sum_{(i,j)\in R} \frac{|V_i||V_j|}{|V|^2} > (1 - \epsilon)$.

We will thus prove the above nonstandard equivalent of the Szemerédi Regularity Lemma. Fix $\epsilon \in \mathbb{R}^{>0}$ and hyperfinite graph (V, E). We consider *three different probability spaces*:

- The Loeb measure space $(V, \Omega_{V,L}, \mu_{V,L})$;
- The product measure space $(V \times V, \Omega_{V,L} \otimes \Omega_{V,L}, \mu_{V,L} \otimes \mu_{V,L})$;
- The Loeb measure space $(V \times V, \Omega_{V\times V,L}, \mu_{V\times V,L})$.

For ease of notation, we will write Ω_V for $\Omega_{V,L}$, $\Omega_{V\times V}$ for $\Omega_{V\times V,L}$, μ_V for $\mu_{V,L}$, and $\mu_{V\times V}$ for $\mu_{V\times V,L}$.

As mentioned in the previous section, $\Omega_V \otimes \Omega_V$ is a σ-subalgebra of $\Omega_{V\times V}$ and $\mu_V \otimes \mu_V$ is the restriction of $\mu_{V\times V}$ to $\Omega_V \otimes \Omega_V$. Since $E \subseteq V \times V$, we can consider $f := \mathbb{E}[1_E|\Omega_V \otimes \Omega_V]$. The following calculation will prove useful: Suppose that $A, B \subseteq V$ are internal and $\frac{|A|}{|V|}$ and $\frac{|B|}{|V|}$ are noninfinitesimal. Then (†):

$$\begin{aligned}
\int_{A\times B} f d(\mu_V \otimes \mu_V) &= \int_{A\times B} 1_E d\mu_{V\times V} \quad \text{by the definition of } f \\
&= \mathrm{st}(\frac{|E \cap (A \times B)|}{|V|^2}) \\
&= \mathrm{st}(\frac{|E \cap (A \times B)|}{|A||B|}) \, \mathrm{st}(\frac{|A||B|}{|V|^2}) \\
&= \mathrm{st}(\delta(A, B)) \, \mathrm{st}(\frac{|A||B|}{|V|^2}).
\end{aligned}$$

Fix $\eta \in \mathbb{R}^{>0}$, to be determined later. Now, since f is $\mu_V \otimes \mu_V$-integrable, there is a $\mu_V \otimes \mu_V$-simple function $g \leq f$ such that $\int (f - g) d(\mu_V \otimes \mu_V) < \eta$. Set $C := \{\omega \in V \times V \mid f(\omega) - g(\omega) \geq \sqrt{\eta}\} \in \Omega_V \otimes \Omega_V$. Then $(\mu_V \otimes \mu_V)(C) < \sqrt{\eta}$, for otherwise

$$\int (f-g)d(\mu_V \otimes \mu_V) \geq \int_C (f-g)d(\mu_V \otimes \mu_V) \geq \int_C \sqrt{\eta} d(\mu_V \otimes \mu_V) \geq \sqrt{\eta}\sqrt{\eta} = \eta.$$

By Fact 15.12, there is a set $D \in \Omega_V \otimes \Omega_V$ that is a finite, disjoint union of rectangles of the form $V' \times V''$, with $V', V'' \subseteq V$ internal sets, such that $C \subseteq D$ and $(\mu_V \otimes \mu_V)(D) < \sqrt{\eta}$. In a similar way, we may assume that the level sets of g (that is, the sets on which g takes constant values) are finite disjoint unions of rectangles (Exercise). We now take a finite partition $V_1, \ldots, V_m$ of V into internal sets such that g and 1_D are constant on each rectangle $V_i \times V_j$. For ease of notation, set d_{ij} to be the constant value of g on $V_i \times V_j$.

Claim: If $\mu_V(V_i), \mu_V(V_j) \neq 0$ and $(V_i \times V_j) \cap D = \emptyset$, then V_i and V_j are internally $2\sqrt{\eta}$-pseudorandom.

Proof of Claim: Since $C \subseteq D$, we have that $(V_i \times V_j) \cap C = \emptyset$, whence

$$d_{ij} \leq f(\omega) < d_{ij} + \sqrt{\eta} \text{ for } \omega \in V_i \times V_j. \quad (\dagger\dagger).$$

Now suppose that $A \subseteq V_i$ and $B \subseteq V_j$ are such that $|A| \geq 2\sqrt{\eta}|V_i|$ and $|B| \geq 2\sqrt{\eta}|V_j|$. In particular, $\frac{|A|}{|V_i|}$ and $\frac{|B|}{|V_j|}$ are noninfinitesimal. Since $\mu_V(V_i), \mu_V(V_j) > 0$, it follows that $\frac{|A|}{|V|}$ and $\frac{|B|}{|V|}$ are noninfinitesimal and the calculation $(\dagger)$ applies. Integrating the inequalities $(\dagger\dagger)$ on $A \times B$ yields:

$$d_{ij} \operatorname{st}(\frac{|A||B|}{|V|^2}) \leq \operatorname{st}(\delta(A,B)) \operatorname{st}(\frac{|A||B|}{|V|^2}) < (d_{ij} + \sqrt{\eta}) \operatorname{st}(\frac{|A||B|}{|V|^2}).$$

We thus get:

$$|\delta(A,B) - \delta(V_i,V_j)| \leq |\delta(A,B) - d_{ij}| + |\delta(V_i,V_j) - d_{ij}| < 2\sqrt{\eta}.$$

By the Claim, we see that we should choose $\eta < (\frac{\epsilon}{2})^2$, so V_i and V_j are internally ϵ-pseudorandom when V_i and V_j are non-null and satisfy $(V_i \times V_j) \cap D = \emptyset$. It remains to observe that the ϵ-pseudorandom pairs almost cover all pairs of vertices. Let $R := \{(i,j) \mid V_i \text{ and } V_j \text{ are } \epsilon\text{-pseudorandom}\}$. Then

$$\begin{aligned}
\operatorname{st}(\sum_{(i,j)\in R} \frac{|V_i||V_j|}{|V|^2}) &= \mu_{V\times V}(\bigcup_{(i,j)\in R} (V_i \times V_j)) \\
&\geq \mu_{V\times V}((V \times V) \setminus D) \\
&> 1 - \sqrt{\eta} \\
&> 1 - \epsilon.
\end{aligned}$$

□

16.1. **Problems.** The Szemerédi Regularity Lemma proven above was integral in Szemerédi's 1975 proof of the following theorem (henceforth referred to as *Szemerédi's Theorem*), which was originally conjectured by Erdős and Turán in 1936:

Theorem 16.3. *If $A \subseteq \mathbb{Z}$ has positive upper Banach density (to be defined below), then A contains arbitrary long nontrivial arithmetic progressions, that is, for every $k \in \mathbb{N}^{>0}$, there is $n, m \in \mathbb{Z}$ with $m > 0$ such that $n, n+m, \ldots, n+(k-1)m \in A$.*

In 1977, Furstenberg gave a different proof of the Szemerédi's Theorem. Rather than using difficult combinatorics a la Szemerédi, Furstenberg translated the question into a question about *ergodic theory* (roughly defined as the study of measure preserving transformations of a probabilty space) and then proved the ergodic version of Szemerdi's Theorem. This translation is now known as the *Furstenberg Correspondence Principle.* In these problems, we will outline a nonstandard proof of the Furtstenburg Correspondence Principle. First, some terminology:

Definition 16.4. If $A \subseteq \mathbb{Z}$, the *upper Banach density of* A is

$$d(A) = \limsup_{m-n\to\infty} \frac{|[n,m]\cap A|}{m-n}.$$

Definition 16.5. If $(X, \mathcal{B}, \mu)$ is a probability space, we say that $T : X \to X$ is a *measure-preserving transformation* if, for all $A \in \mathcal{B}$, both $T(A)$ and $T^{-1}(A)$ belong to $\mathcal{B}$ and $\mu(T(A)) = \mu(T^{-1}(A)) = \mu(A)$.

Here is the theorem:

Theorem 16.6 (Furstenburg Correspondence Principle)**.** *Suppose that $A \subseteq \mathbb{Z}$ has positive upper Banach density. Then there is a probability space $(X, \mathcal{B}, \mu)$, a measure-preserving transformation $T : X \to X$, and a measurable set $A_0 \in \mathcal{B}$ such that $\mu(A_0) = d(A)$ and such that, for any finite set $U \subseteq \mathbb{Z}$, we have:*

$$d(\bigcap_{i\in U}(A-i)) \geq \mu(\bigcap_{i\in U} T^{-i}(A_0)).$$

Before proving the Furstenberg Correspondence Princple, let us mention the ergodic-theoretic fact that Furstenburg proved:

Theorem 16.7. *Suppose that $T : X \to X$ is a measure preserving transformation on the probability space X, $\mu(A) > 0$, and $k \in \mathbb{N}$. Then there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \cdots \cap T^{-(k-1)n}(A)) > 0$.*

Notice that the above theorem, coupled with the Furstenberg Correspondence Principle, yields a proof of the Szemerédi Theorem.

Problem 16.1. If $A \subseteq \mathbb{Z}$, prove that

$$d(A) = \max\{\mathrm{st}(\frac{|A^*\cap I|}{|I|}) \mid I \subseteq \mathbb{Z}^* \text{ is an interval of hyperfinite length}\}.$$

(Here, when we say "hyperfinite," we mean "hyperfinite, but not finite".)

Problem 16.2. Fix $A \subseteq \mathbb{Z}$ with positive upper Banach density. Fix $I \subseteq \mathbb{Z}^*$, an interval of hyperfinite length such that $\frac{|A^*\cap I|}{|I|} \approx d(A)$. Show that:

- the Loeb measure space (I, Ω_I, μ_I),
- the map $T : E \to E$ given by $T(x) = x + 1(\operatorname{mod} I)$, and
- the set $A_0 := A^* \cap I$

satisfy the conclusion of the Furstenberg Correspondence Princple.

References

[1] M. Davis, *Applied nonstandard analysis*, John Wiley and Sons Inc., 1977.

[2] R. Goldblatt, *Lectures on the hyperreals: an introduction to nonstandard analysis*, volume 188 of Graduate Texts in Mathematics. Springer-Verglag, 1998

[3] C. W. Henson, *Foundations of Nonstandard Analysis: A Gentle Introduction to Nonstandard Extensions*; Nonstandard Analysis: Theory and Applications, L.O. Arkeryd, N.J. Cutland, and C.W. Henson, eds., NATO Science Series C; Springer 2001.

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Science and Engineering Offices M/C 249, 851 S. Morgan St., Chicago, IL, 60607-7045

E-mail address: isaac@math.uic.edu

URL: http://www.math.uic.edu/~isaac