An Ultrashort Introduction to Nonstandard Analysis and its Application to Ergodic Theory

— Notes —

Heinrich-Gregor Zirnstein

October 16, 2012

1 About these Notes

These notes grew out of a talk I gave at the Oberwolfach Arbeitsgemeinschaft "Ergodic Theory and Combinatorial Number Theory" in October 2012. Several people expressed a desire for more information and I figured it would be useful to write up my personal notes and add a little bibliography.

The aim is to give a very short introduction to nonstandard analysis and its recent uses in ergodic theory. For instance, structure theorems for ultraproducts of compact groups [Sze12] and constructing measure preserving systems as Loeb measure spaces [Tow09] have recently attracted attention.

In these notes, I will not prove most of the basic facts about nonstandard analysis, but instead use them to present a simple nonstandard proof [Kam82] of the pointwise ergodic theorem. For readers desiring to know more about the technical details of nonstandard analysis, I would recommend the following books:

- [Rob96] is the original referencefor nonstandard analysis and explores some consequences in functional anlysis.
- [Cut04] introduces and uses Loeb measures.
- [LR94] is a comprehensive collection of technical facts.

2 Nonstandard Mathematics

2.1 The intuition of introducing "infinite" numbers

The original motivation for nonstandard mathematics was to make rigorous the notions of *infinitely large* and *infinitesimally small numbers*.

Of course, we know that there is no natural number $n \in \mathbb{N}$ that is larger than all other numbers. (Exercise!) But we can construct a new kind of numbers, the *nonstandard natural numbers* ${}^*\mathbb{N}$. They contain the standard numbers $\mathbb{N} \subseteq {}^*\mathbb{N}$, but they also contain nonstandard numbers that are larger than any standard number.

Adding infinite numbers is somewhat similar to the process of extending number fields in algebra. For instance, we could consider the ring $\mathbb{Z}[\sqrt{2}]$ that arises by adding a formal solution t of the polynomial equation $t^2 - 2 = 0$ to the ring of integers $\mathbb{Z}$. Likewise, we want to add an infinite number α to the set of natural number $\mathbb{N}$. But what equations should this new "number" α satisfy? Clearly, an infinite number does not satisfy any particular polynomial equation $\alpha^3+\alpha+4=0$, because solutions to these equations behave like finite numbers.

A good answer to the question of equations is that an infinite number α should be *"generic"* in some way. It should *not* fulfill any special equations, but it should retain the properties that are satisfied by "most" ordinary numbers anyway. For instance, most natural numbers n are larger than the particular number 7, so the generic number α should also be larger than 7. Here, "most" means "all but finitely many exceptions". This is very nice, because we already see that the "generic" number α is greater than any particular number, so generic numbers behaves like infinite numbers.

Robinson [Rob96] and independently Laugwitz [Lau86] showed how to make this intuition rigorous. The key technical ingredients are *ultrafilters* and the construction of *ultrapowers*.

2.2 Ultrafilters

Ultrafilters are a good way to make rigorous the notion of "most natural numbers".

Definition 2.1. *Let X be a set, for instance $X = \mathbb{N}$. Consider a collection $p \subseteq \mathcal{P}(X)$ of its subsets. This collection is called an* ultrafilter *iff it satisfies the following four conditions*

1. $\emptyset \notin p$
2. $A \in p$ *and* $A \subseteq B \implies B \in p$
3. $A \in p$ *and* $B \in p \implies A \cap B \in p$
4. $A \cup B \in p \implies A \in p$ *or* $B \in p$

I have to admit that I do not understand the meaning of the previous definition at all, at least not as written. What I do understand, however, instead is the following:

Imagine that for some reason, we have completely forgotten what an element of a subset $A \subseteq X$ is, i.e. that we have no memory of what the relation $\in$ means, but that we still remember what subsets are, i.e. we still remember the operations $\emptyset, \subseteq, \cap, \cup$ on subsets. How could we axiomatize the notion of "is an element of" given only our knowledge of subsets? The following axioms for the predicate $\ni$ seem very reasonable

1. $\emptyset \not\ni p$

2. $A \ni p$ and $A \subseteq B \implies B \ni p$
3. $A \ni p$ and $B \ni p \implies A \cap B \ni p$
4. $A \cup B \ni p \implies A \ni p$ or $B \ni p$

But these are precisely the axioms of ultrafilters with the sign $\in$ flipped around to read $\ni$!

In other words, ultrafilters can be understood as "generalized points". An ultrafilter p is identified with the collection of sets that its corresponding "generalized point" is contained in.

Obviously, then, the ordinary points are also generic points. For instance, an ultrafilter p_5 corresonding to the point $5 \in \mathbb{N}$ is given by the collection of sets

$$p_5 = \{A : 5 \in A, A \subseteq \mathbb{N}\}$$

Ultrafilters of this form are called *principal*.

Now, the interesting thing is that there also exist ultrafilters which are *not* of this form, they are called *non-principal* ultrafilters. They represent new generalized points and are hence very interesting. However, their construction requires the Axiom of Choice[1] and we can never write down some explicit representation for them.

Exercise. Start with the collection of sets

$$\mathcal{F} := \{A : A \subseteq \mathbb{N},\ \text{there exists } n \in \mathbb{N} \text{ such that } \{n, n+1, n+2, \dots\} \subset A\}$$

and convince yourself that it satisfies the first three axioms of an ultrafilter, but not the fourth. A collection with this property is called a *filter*. The particular filter above is called the *Fréchet filter*. Use the Axiom of Choice to add new subsets to it until you get a collection $p \supset \mathcal{F}$ that satisfies the fourth axiom as well, i.e. extend this filter $\mathcal{F}$ to an ultrafilter p. Observe that this ultrafilter is non-principal.

It turns out that all non-principal ultrafilters are extensions of the Fréchet filter. Thanks to this and the fourth axiom, they implement very well the notion of "most": given a non-principal ultrafilter p, we can say that a set $A \subseteq \mathbb{N}$ contains "most" numbers if and only if $A \in p$.

[1] Actually, the existence of ultrafilters is not equivalent to the Axiom of Choice, but it is still outside the ZF axioms of set theory.

2.3 Basic notions of nonstandard analysis

Using ultrafilters, we can now construct the so-called *ultrapowers*, which will give us infinite numbers as mentioned in the introduction.

Definition 2.2. *Pick a non-principal ultrafilter p. Consider a set V, for instance the set of natural numbers $V = \mathbb{N}$. The* ultrapower *${}^*V$ of the set V is defined as the set of equivalence classes of sequences*

$$ {}^*V := \prod_{n \to p} V := (\prod_{n \in \mathbb{N}} V)/\sim $$

where two sequences are considered equal if they are equal for "most" indices

$$ (a_n) \sim (b_n) \iff \{n : a_n = b_n\} \in p. $$

Remark. The constant sequences embed $V \subseteq {}^*V$. These elements are called *standard.*

Arithmetical operations are extended element-wise to the nonstandard space

Definition 2.3. *For $a, b \in {}^*\mathbb{N}$, define $a + b \in {}^*\mathbb{N}$ via*

$$ (a + b)_n := a_n + b_n. $$

For comparison operations, define $a < b \in {}^*\mathbb{N}$ via

$$ a < b := \{n : a_n < b_n, n \in \mathbb{N}\} \in p $$

and so on.

The axioms of ultrafilters guarantee that this is well-defined with respect to the equivalence relation $\sim$. Likewise for all other operations, comparisons, etc.

The first example of an element of ${}^*V$ that is not in V is given by the (equivalence class of the) sequence $c_n = n$.

Definition 2.4. *A number $c \in {}^*\mathbb{N}$ that is larger than any standard number*

$$ \forall m \in \mathbb{N}.\ c > m $$

is called hyperfinite.

Example: $c_n = n$. This is where we need that the ultrafilter p is non-principal.

In many respects, however, the ultrapower ${}^*V$ behaves exactly like the original set V. Most importantly, they fulfill the same logical *first-order formulas.* Example: the logical formula

$$\forall n \in \mathbb{N}.\exists m \in \mathbb{N}.\ n = 2m + 1 \vee n = 2m$$

is true exactly when the formula

$$\forall n \in {}^*\mathbb{N}.\exists m \in {}^*\mathbb{N}.\ n = 2m + 1 \vee n = 2m$$

is true. A first-order formula contains only quantifiers over the set V, i.e. $\forall v \in V$, $\exists v \in V$. In particular, all the usual properties for addition $+$, multiplication and so on continue to hold.

Not all formulas are first order. For example, consider the induction principle: every set of natural numbers has a smallest elements. This is a second-order formula as it quantifies over subsets of the natural numbers.

$$\forall .S \in \mathcal{P}(\mathbb{N}).\exists s \in \mathbb{N}.\forall x \in \mathbb{N}.s \in S \wedge (x \in S \implies s \leq x).$$

Even in this case, we get a nonstandard version.

$$\forall .S \in {}^*\mathcal{P}(\mathbb{N}).\exists s \in {}^*\mathbb{N}.\forall x \in {}^*\mathbb{N}.s \in S \wedge (x \in S \implies s \leq x)$$

This procedure works for all formulas. I don't want to formalize this in detail, but you get the idea.

Theorem 2.5 (Transfer principle)**.** *Take a logical formula that is true and replace* every *quantification over $V, \mathcal{P}(V)$ and so on with a quantification over the corresponding ultrapower ${}^*V, {}^*\mathcal{P}(V), \ldots$. The resulting formula will also be true.*

It is important to replace *every* quantification.

Crucially, and this is where things become interesting, we now have two different notions of powersets

$${}^*\mathcal{P}(B) \subsetneq \mathcal{P}({}^*V).$$

Definition 2.6. *Elements $S \in {}^*\mathcal{P}(V)$ are called* internal sets. *They can be identified with sets of nonstandard elements by extending the $\in$ relation to the ultrapower*

$$x \in S :\iff \{n : x_n \in S_n\} \in p.$$

All other sets of nonstandard elements are called external.

Definition 2.7. *Likewise, elements $f \in {}^*(V \to W)$ are called* internal functions. *They can be identified with ordinary functions between ultrapowers.*

Due to the transfer principle, internal sets behave very much like finite sets. For instance, we have

Proposition 2.8. *Every internal set has a smallest element.*

Corollary 2.9. *The set of hyperfinite ("infinitely large") numbers is external.*

Corollary 2.10. *The set of finite numbers* $\mathbb{N}$ *is external.*

Corollary 2.11 (Underflow)**.** *Let* $S \subset {}^*\mathbb{N}$ *be an internal set that contains all hyperfinite numbers. Then it also contains a standard number,* $\mathbb{N} \cap S \neq \emptyset$*.*

We also want to introduce the notion of infinitesimal numbers.

Definition 2.12. *Let* $r \in {}^*\mathbb{R}$ *be a hyperreal number.*

- *The hyperreal number* r *is called* finite *if it can be bounded by a standard number*

$$\exists L \in \mathbb{R}. |r| < L.$$

- *The hyperreal number* r *is called* infinitesimal *if it is smaller than every standard number*

$$\forall \varepsilon \in \mathbb{R}.\ |r| < \varepsilon.$$

- *Every finite number has a* standard part

$$st(r) := \text{ unique } x \in \mathbb{R} \text{ such that } x - r \text{ is infinitesimal}$$

2.4 Loeb measure

Having infinite numbers at our disposal, we can now make sense of infinite sums and integrals. In the nonstandard world, integrals *are* infinite sums.

In particular, let c be a hyperfinite number and $[1, c] \subset {}^*\mathbb{N}$ the corresponding hyperfinite interval of numbers up to c.

Theorem 2.13. *(Loeb measure) Consider a hyperfinite interval* $[1, c]$*. There exists a* σ*-algebra* $\mathcal{L}$ *and a unique measure* ν *on* $\mathcal{L}$*, the* Loeb measure*, with the following properties*

- *The* σ*-algebra* $\mathcal{L}$ *contains all internal subsets of this hyperfinite interval.*
- *The Loeb measure counts elements*

$$\nu(A) = st\left(\frac{|A|}{c}\right) \qquad \text{for } A \text{ internal}$$

- *The σ-algebra $\mathcal{L}$ is complete in the measure-theoretic sense*

$$A \subseteq B, B \in \mathcal{L}, \nu(B) = 0 \implies A \in \mathcal{L}$$

The construction of the Loeb measure can be done with the Carathéodory extension theorem in a way similar to the construction of the Lebesgue measure. We skip the proofs, but note the following points:

The σ-additivity of the content ν actually follows from additivity, because of the following curious compactness property:

Lemma 2.14. *Let B_n be a sequence of internal sets such that the countable union $\cup_{n\in\mathbb{N}} B_n = B$ is again an internal set. Then, it is already a* finite *union $\cup_{n\leq m} B_n = B$.*

However, the collection of internal sets is not a σ-algebra. After all, the set $\mathbb{N}$ is not internal. Still, they almost form a σ-algebra, in the sense of doing that up to sets of measure zero.

Lemma 2.15. *Let X be a Loeb-measurable set. Then, there exists an internal set A with the property*

$$\nu(X\Delta A) = 0.$$

Remember that in the case of the Lebesgue measure, we usually incur an error of $\varepsilon > 0$ when approximating measurable sets by a collection of basic intervals.

Now, given the Loeb measure, we can define measurable and integrable functions as usual. Internal functions are measurable and have a very simple integral.

Lemma 2.16. *Let $f : [1, c] \to {}^*\mathbb{R}$ be a finitely bounded internal function. Then, the function $st \circ f$ is integrable and we have*

$$\int st(f(x))d\nu(x) = st\left(\frac{1}{c}\sum_{k=1}^{c} f(k)\right).$$

A general integrable function can be approximated by internal functions.

Lemma 2.17. *Let $f : [1, c] \to \mathbb{R}$ be a Loeb integrable function. For any standard $\varepsilon > 0$, there exist internal functions $F, G : [1, c] \to {}^*\mathbb{R}$ such that*

- *$F(x) \leq f(x) \leq G(x)$ for all $x \in [1, c]$.*
- *$|\int_B f d\nu - \frac{1}{c}\sum_{k\in B} F(k)| < \varepsilon$ and $|\int_B f d\nu - \frac{1}{c}\sum_{k\in B} G(k)| < \varepsilon$ for all internal sets B.*

3 The Pointwise Ergodic Theorem

We now want to present a simple proof of the ergodic theorem using nonstandard analysis. [Kam82]

Theorem 3.1 (Pointwise ergodic theorem)**.** *Let* $(X, \mathcal{X}, \mu, T)$ *be a measure preserving system. For any function* $f \in L^1(X, \mathcal{X}, \mu)$*, the limit*

$$\hat{f}(x) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(T^k x)$$

exists for almost all x *and is integrable with*

$$\int \hat{f} d\nu = \int f d\nu.$$

The nonstandard proof can actually be turned into an equally simple standard proof, but the author would like to present the nonstandard version, as it teaches us something about the construction of measure spaces via nonstandard analysis and features a neat compactness argument.

In fact, it will suffice to prove the pointwise ergodic theorem for the following particular dynamic system

Definition 3.2. *Consider a hyperfinite interval* $[1, c]$ *and the measure preserving system* $([1, c], \mathcal{L}, \nu, T)$ *with the transformation*

$$T(n) = (n + 1) \mod c.$$

This system is called a universal system*.*

As we will see in a moment, this is the grandfather of all measure preserving systems. Well not all of them, but at least of those on a standard Borel space. Of course, grandpa turns out to be a bit of a hypocrite: his own measure space is *not* a standard Borel space.

Lemma 3.3. *Every standard Borel space* $([0, 1], \mathcal{B}, \lambda)$ *with a measure preserving transformation* T *is a factor of the universal system.*

Proof. (Sketch.)

Consider for a moment a single point $x_0 \in X$. If the transformation were to send this point all across the space in an evenly distributed manner, we could evaluate the integral of a continuous function $f : X \to \mathbb{R}$ by sampling

$$\int f(x)dx \approx \frac{1}{N}\sum_{k=1}^{N} f(T^k x_0).$$

The right-hand side corresponds to the Loeb measure on the universal system, where we map each integer k to the point $T^k x_0$. This is called a *typical* point x_0.

Of course, no such point x_0 needs to exist, for instance because the transformation T does not "mix" the sample space enough. But we can consider a *pair* (x_0, y_0) of points instead that trace out different parts of the sample space. More generally, if we take a large enough ensemble of points, they will trace out the sample space.

Technically this can be achieved by considering the isomorphic system of orbits

$$([0,1], \mathcal{B}, \lambda, T) \tilde{\to} ([0,1]^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mu, \sigma)$$

$$x \mapsto (T^n x)_{n\in\mathbb{N}}$$

where σ is the standard shift and μ the push-forward measure. Unlike the former system, the latter system contains a typical point, because it also contains sequences that are not orbits of a single point, but instead pieced together from orbits of many different points. Details can be found in [Kam82]. ☐

Fortunately, the general case of the pointwise ergodic theorem can be reduced to the case of standard Borel spaces.

Lemma 3.4. *If the pointwise ergodic theorem holds for standard Borel spaces, then it holds for all measure preserving systems.*

Proof. (Exercise) If we interpreter the measure space as a probability space, we don't really care about its elements, but only about the values of the random variables

$$x \mapsto (f(T^n x))_{n\in\mathbb{N}}$$

This map is a factor map onto a standard Borel space. ☐

With these preparations, we can finally proceed to prove the pointwise ergodic theorem.

Proof. (Pointwise ergodic theorem) By the previous arguments, it is enough to show this for the universal system $L^1([1,c], \mathcal{L}, \nu)$.

Consider the pointwise upper limit

$$\bar{f}(x) = \limsup_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x)$$

and the pointwise lower limit

$$\underline{f}(x) = \liminf_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x).$$

It is sufficient to show that

$$\int \bar{f} d\nu \leq \int f d\nu \leq \int \underline{f} d\nu.$$

For simplicity, we only consider the first inequality, the second is analogous. Moreover, we assume that the function f is bounded as $0 \leq f \leq L$; lifting these restrictions is an easy exercise.

Let us fix $\varepsilon > 0$ and approximate the two functions by internal functions

$$G(x) \leq \bar{f}(x), \quad f(x) \leq F(x), \quad F, G \text{ internal}$$

$$|\int G d\nu - \int f d\nu| < \varepsilon, \quad |\int F d\nu - \int \bar{f} d\nu| < \varepsilon.$$

Now, due to the nature of the upper limit, for each point x, there exists a natural number m such that the average comes close to its upper limit point

$$\bar{f}(x) - \varepsilon \leq \frac{1}{m} \sum_{k=0}^{m-1} f(T^k x).$$

In this inequality, we replace $f, \bar{f}$ by their internal approximations F, G and consider the smallest integer m such that

$$m(x) = \min\{m \in \mathbb{N} : \bar{G}(x) - \varepsilon \leq \frac{1}{m} \sum_{k=0}^{m-1} F(T^k x)\} \in \mathbb{N}$$

This is a finite integer for each point x of the hyperfinite interval. But here's the kicker: the function $m(x)$ is an *internal* function and this means that we have a finite global bound $m(x) \leq M \in \mathbb{N}$!

In other words, if we now consider the sequence of $N \in \mathbb{N}$ function values

$$F(T^1 y), F(T^2 y), F(T^3 y), \ldots, F(T^N y)$$

we have to take the average of the first $m(y)$ sequence elements to be close to the upper limit of the total average. But these are at most M elements. Considering the average of the next $m(T^{m(y)+1}y)$ elements and so on, we see that a very long average can be split into several parts with close averages and a leftover part of length at most M. This means

$$\frac{N-M}{N}(G(x) - \varepsilon) \leq \frac{1}{N}\sum_{k=0}^{N-1} F(T^k x).$$

Taking the integral on both sides, letting $N \to \infty$ and remembering the original functions, we see that

$$\int \bar{f} d\nu \leq \int G d\nu + \varepsilon \leq \int F d\nu + 2\varepsilon \leq \int f d\nu + 3\varepsilon.$$

□

Remark. It is not difficult to recast the preceding proof entirely as a standard argument. Essentially, the most important change needed is that the global bound M will only hold on a T-invariant set whose complement has small measure. However, I wanted to demonstrate the use of Loeb spaces as universal measure preserving systems, and I wanted to share the nice argument where we obtained a global upper bound for an internal function.

References

[Cut04] Nigel J Cutland. *Loeb measures in practice: recent advances*. Vol. 1751. Lecture Notes in Mathematics. 2004. DOI: 10.1007/b76881.

[Kam82] Teturo Kamae. “A simple proof of the ergodic theorem using nonstandard analysis”. In: *Israel Journal of Mathematics* 42.4 (1982), pp. 284–290. URL: http://www.springerlink.com/content/m17vg3613763603m/?MUD=MP.

[Lau86] Detlef Laugwitz. *Zahlen und Kontinuum: eine Einführung in die Infinitesimalmathematik*. German. Vol. 5. Lehrbücher und Monographien zur Didaktik der Mathematik. Bibliographisches Institut, 1986. URL: http://books.google.de/books?id=1xzvAAAAMAAJ.

[LR94] Dieter Landers and Lothar Rogge. *Nichtstandard Analysis*. German. Springer Verlag, Mar. 1994. ISBN: 9783540571155. URL: http://books.google.de/books?id=XGIp68yCPjcC.

[Rob96] Abraham Robinson. *Non-standard analysis*. Princeton University Press, 1996. ISBN: 9780691044903. URL: http://books.google.de/books?id=OkONWa4ToH4C.

[Sze12] Balázs Szegedy. “On higher order Fourier analysis”. In: *arXiv.org* math.CO (Mar. 2012). URL: http://arxiv.org/abs/1203.2260.

[Tow09] Henry Towsner. “Convergence of Diagonal Ergodic Averages”. In: *Ergodic Theory and Dynamical Systems* 29.4 (Aug. 2009), pp. 1309–1326. URL: http://arxiv.org/abs/0711.1180.