

**Aspects of Ergodic Theory  
in  
Subsystems  
of  
Second-order Arithmetic**

*Ksenija Simic*  
Carnegie Mellon University  
*Thesis advisor: Jeremy Avigad*

Submitted to  
The Department of Mathematical Sciences,  
in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

May 5, 2004



*The ordinary-sized stuff which is our lives, the things people write poetry about - clouds - daffodils - waterfalls - and what happens in a cup of coffee when the cream goes in - these things are full of mystery, as mysterious to us as the heavens were to the Greeks. We're better at predicting events at the edge of the galaxy or inside the nucleus of an atom than whether it'll rain on auntie's garden party three Sundays from now. Because the problem turns out to be different. We can't even predict the next drip from a dripping tap when it gets irregular. Each drip sets up conditions for the next, the smallest variation blows predictions apart, and the weather is unpredictable the same way, will always be unpredictable. When you push numbers through the computer you can see it on the screen. The future is disorder. A door like this has cracked open five or six times since we got up on our hind legs. It's the best possible time to be alive, when almost everything you thought you knew is wrong.*

(Tom Stoppard, *Arcadia*)



# Acknowledgements

First and foremost, I wish to thank my advisor Jeremy Avigad who, throughout the time we worked together never ran out of ideas, patience, or words of encouragement, and whose passion for mathematics always inspires me. Many of the results in the dissertation came out of our joint work. I am also grateful to my thesis committee members: James Cummings, William Hrusa, Richard Statman and Stephen Simpson, for careful reading of my work and their suggestions, as well as to the mathematics department, in particular Russ Walker and Bill Hrusa for their support, as well as to Stella, Ferna and Nancy for always being extraordinarily helpful and kind.

I am grateful (beyond words) to my parents and sister as well as to my oldest friends from the old country who never for a moment entertained the notion that I could fail at anything I do.

Thanks to Chad for helping me through the first two years and to my office-mates from the math ghetto also known as the physical plant building for friendship and great conversation, to Linda Hooper for endless emotional support, to my students for making teaching great, to Dr. Hirsch, and all the other wonderful people of Pittsburgh and the world who made the last six years not so bad at all.



# Contents

|   |            |
|---|------------|
| <b>Acknowledgements</b>                               | <b>iii</b> |
| <b>Introduction</b>                                   | <b>vii</b> |
| <b>1 General Preliminaries</b>                        | <b>1</b>   |
| 1.1 Definitions . . . . .                             | 1          |
| 1.2 Some Useful Facts . . . . .                       | 6          |
| 1.3 Brief Overview of Ergodic Theory . . . . .        | 7          |
| 1.4 Challenges of Reverse Mathematics . . . . .       | 9          |
| 1.5 Parallels With Constructive Mathematics . . . . . | 11         |
| <b>2 Banach and Hilbert Spaces</b>                    | <b>15</b>  |
| 2.1 Preliminaries . . . . .                           | 15         |
| 2.2 Properties of Banach and Hilbert Spaces . . . . . | 19         |
| 2.3 Bounded Linear Functionals . . . . .              | 28         |
| <b>3 The Mean Ergodic Theorem</b>                     | <b>33</b>  |
| 3.1 Proof of the Main Theorem . . . . .               | 33         |
| 3.2 Reversal . . . . .                                | 37         |
| <b>4 Aspects of Measure Theory</b>                    | <b>41</b>  |
| 4.1 Preliminaries . . . . .                           | 41         |
| 4.2 Products and Powers in $L_p$ Spaces . . . . .     | 48         |
| 4.3 Further Properties of $L_p$ Spaces . . . . .      | 56         |
| 4.4 Integrable Functions and Sets . . . . .           | 60         |
| 4.5 The Monotone Convergence Theorem . . . . .        | 64         |
| <b>5 The Pointwise Ergodic Theorem</b>                | <b>69</b>  |
| 5.1 Motivation . . . . .                              | 70         |
| 5.2 Convergence in Norm . . . . .                     | 76         |

|          |   |            |
|----------|---|------------|
| 5.3      | The Maximal Ergodic Theorem . . . . .     | 80         |
| 5.4      | First Proof of the Main Theorem . . . . . | 82         |
| 5.5      | Second Proof . . . . .                    | 84         |
| 5.6      | Third Proof . . . . .                     | 88         |
| 5.7      | Reversal . . . . .                        | 93         |
| <b>6</b> | <b>Closing Remarks</b>                    | <b>99</b>  |
|          | <b>Appendix A</b>                         | <b>101</b> |
|          | <b>Appendix B</b>                         | <b>103</b> |



# Introduction

Formalizing mathematics within second-order arithmetic is far from being a novel venture. On the contrary, the origin of this practice can be traced back to the beginning of the twentieth century. The turn of the last century saw major changes in both practice and philosophy of mathematics, one of the main new developments being the introduction of formal methods. The greatest mathematicians of the time, including Dedekind, Frege, Peano, Russell and Weyl, contributed, in different ways, to the development of axiomatic foundations for mathematics. It would be inaccurate to say that these mathematicians began the development of mathematics within second-order arithmetic, since for the most part they worked in higher-order logic or without an underlying formal system. It wasn't until Hilbert that first and second-order logic were clearly separated from higher-order logic. It was also Hilbert who first advocated second-order arithmetic as a formal system suitable for development of analysis. He wrote about first and second-order logic as early as 1917/18 in his lecture notes for the course "Prinzipien der Mathematik," but the true pioneering work in this area was *Grundlagen der Mathematik* that Hilbert wrote with Bernays, and which was first published in 1934. More than just recognizing second-order logic as a separate entity, they gave axioms for second-order arithmetic and expressed the belief that analysis can be done naturally in this formal system. They proceeded to develop some of the theory and proved a number of results. Between the 1950's and 1970's this subject was taken up by a number of mathematicians, including Kleene, Kreisel, Friedman, Feferman and Takeuti, who undertook an even finer analysis, dividing second-order arithmetic into subsystems of varying logical strengths. The question arose of which of the standard mathematical theorems are provable in which fragments of second-order arithmetic, and has since been successfully answered in a number of different areas. More on this subject, in the broader context of historical development of proof theory, can be found in [1].

A question that often seems to arise in connection to doing mathematics

in second-order mathematics is: what is the purpose of this endeavor? If set theory provides a sound foundation for the practice of mathematics, why forsake it for a much weaker theory? The simple answer is this: using the complex machinery of ZFC for proving theorems of countable mathematics is, in a sense, an overkill. It is frequently the case that only a fragment of full set theory is used in a proof. Our goal is to determine exactly what this fragment is. We want to know not only whether a theorem can be formalized and proved in second-order arithmetic, but also precisely how much of its full strength is used. The motivation for this approach is primarily foundational, though it should be noted that it also provides insight into the structure and complexity of proofs and the mathematical concepts involved.

It is true that the language of second-order arithmetic is limited: the only objects it allows for are natural numbers and sets of natural numbers. But if we focus our attention on non set-theoretic mathematics, it turns out that second-order arithmetic suffices for most considerations, both with respect to representations and proofs.

Full second-order arithmetic consists of axioms for an ordered semiring, the induction axiom and the comprehension scheme

$$\exists X \forall n (n \in X \leftrightarrow \varphi(n))$$

where  $\varphi$  is any formula of the language in which  $X$  doesn't occur freely. This implies that every set defined by a formula in the language exists. In most cases, however, full comprehension is more than is needed to prove a theorem and can be replaced with weaker set existence axioms, yielding different subsystems of second-order arithmetic. Conveniently, there is no need for a whole array of such subsystems: six suffice to prove most theorems of countable mathematics, and I will for the most part be using only three out of those six.

The name frequently used for this area of research is *reverse mathematics*. It comes from the nature of the pursuit: not only do the axioms apply to certain theorems, but oftentimes the reverse is also true, and most relevant theorems are logically equivalent to set existence axioms used to prove them. For example, the Heine-Borel covering theorem is equivalent to weak König's lemma over the base system  $\text{RCA}_0$  and the Bolzano-Weierstrass theorem is equivalent to arithmetic comprehension (*ACA*) over  $\text{RCA}_0$ . (The subsystems will be defined in Chapter 1.)

Principal investigators in this area in the past twenty years or so have been Harvey Friedman, Stephen Simpson and Simpson's students. Large portions of mathematics have been formalized, including topology, measure

theory, algebra. My research builds on this work, especially that in measure theory. At the same time, it deals with a number of topics that were not previously addressed, such as the theory of Hilbert spaces. Although the title of the thesis indicates that it is an investigation into ergodic theory, proofs of the mean and pointwise ergodic theorems are only the end result of a broader work. Substantial effort needs to be put into setting the grounds for proving them.

## Outline

Chapter 1 provides general definitions and preliminaries, as well as a brief overview of ergodic theory, a discussion of some of the peculiarities and possible drawbacks of working in second-order arithmetic, and a comparison with constructive mathematics. Hilbert spaces are discussed in the second chapter, in particular, existence of orthogonal projections and properties of bounded linear functionals, interesting in their own right, but essential in the proof of the mean ergodic theorem. Next follows the proof of this theorem and its reversal to  $(ACA)$  for a general Hilbert space. Then, before moving on to the pointwise ergodic theorem, an entire chapter (Chapter 4) is dedicated to measure theory. Though it is almost entirely self-contained, it lists properties of integrable functions (and more generally elements of  $L_p$  spaces) needed to prove the pointwise ergodic theorem. The last chapter contains three proofs of the pointwise ergodic theorem and its reversal to  $(ACA)$ . The appendices contain proofs of a few facts that are used in the text, but are not necessary for its understanding.



# Chapter 1

## General Preliminaries

### 1.1 Definitions

*Note:* Most definitions in this and other chapters are contained in Simpson's monograph [26], which is the most comprehensive account on the topic of second-order arithmetic to date.

The language of second-order arithmetic (usually referred to as  $L_2$ ) is a two-sorted language. This means that there are two types of variables: *number variables*, denoted as  $n, m, \dots$  and intended to range over natural numbers, and *set variables*, denoted as  $X, Y, \dots$  intended to range over sets of natural numbers. Terms are built from variables and constant symbols 0 and 1. Atomic formulas are of the form  $t_1 = t_2$ ,  $t_1 < t_2$  and  $t_1 \in X$ , where  $t_1$  and  $t_2$  are terms. Finally, formulas are built from atomic formulas by means of propositional connectives, and number and set quantifiers.

Full second-order arithmetic (usually referred to as  $Z_2$  or  $\Pi_\infty^1\text{-CA}_0$ ) is formalized by classical propositional logic and axioms for second-order arithmetic. The latter are divided into:

- Basic (ordered semiring) axioms:

$$n + 1 \neq 0$$

$$m + 1 = n + 1 \rightarrow m = n$$

$$m + 0 = m$$

$$m + (n + 1) = (m + n) + 1$$

$$m \cdot 0 = 0$$

$$m \cdot (n + 1) = (m \cdot n) + m$$

$$\neg(m < 0)$$

$$m < n + 1 \leftrightarrow (m < n \vee m = n);$$

- Induction axiom

$$(0 \in X \wedge \forall n (n \in X \rightarrow n + 1 \in X)) \rightarrow \forall n (n \in X);$$

- Comprehension scheme

$$\exists X (\varphi(n) \leftrightarrow n \in X),$$

where  $\varphi(n)$  is any formula in  $L_2$  and  $X$  is not free in  $\varphi(n)$ .

Different subsystems of second-order arithmetic are obtained by restricting comprehension and replacing it with weaker set existence axioms. In some cases induction is also modified.

**Definition 1.1.1** *An  $L_2$  formula is said to be a bounded quantifier formula if all the quantifiers occurring in it are of the form  $\forall n < t$ ,  $\exists n < t$ ,  $\forall n \leq t$  or  $\exists n \leq t$ .*

*An  $L_2$  formula is  $\Sigma_1^0$  if it is of the form  $\exists n \varphi$  where  $n$  is a number variable and  $\varphi$  a bounded quantifier formula. Similarly, a  $\Pi_1^0$  formula is of the form  $\forall n \varphi$ .*

*More generally, a  $\Sigma_k^0$  formula is of the form  $\exists n_1 \forall n_2 \exists n_3 \dots n_k \varphi$ , where  $n_1, \dots, n_k$  are number variables and  $\varphi$  is a bounded quantifier formula. Similarly, a  $\Pi_k^0$  formula is of the form  $\forall n_1 \exists n_2 \dots n_k \varphi$ .*

*A formula of  $L_2$  is an arithmetical formula if it is equivalent to a  $\Sigma_k^0$  or  $\Pi_k^0$  formula for some  $k$ .*

*A formula is  $\Pi_1^1$  if it is of the form  $\forall X \theta$ , where  $X$  is a set variable, and  $\theta$  is an arithmetic formula.*

*Note:* If a formula being considered has only first-order quantifiers, and if there is no confusion, the superscript 0 will be omitted. For example,  $\Sigma_1$  stands for  $\Sigma_1^0$  etc.

Most countable mathematics can be formalized in one of the five systems listed below, ordered from weakest to strongest. The subscript 0 will stand for restricted induction, since no system allows induction over an arbitrary formula in the language.

- $[RCA_0]$ . Basic axioms, plus  $\Sigma_1$  induction and  $\Delta_1$  comprehension.
- $[WKL_0]$ .  $RCA_0$  together with weak König's Lemma ( $WKL$ ), a compactness principle stating that every infinite subtree of  $2^{<\mathbb{N}}$  has a path.
- $[ACA_0]$ .  $RCA_0$  plus arithmetic comprehension axiom ( $ACA$ ).

- [ATR<sub>0</sub>]. ACA<sub>0</sub> with arithmetical transfinite recursion.
- [Π<sub>1</sub><sup>1</sup>-CA<sub>0</sub>]. ACA<sub>0</sub> plus Π<sub>1</sub><sup>1</sup> comprehension (Π<sub>1</sub><sup>1</sup>-CA).

Though Simpson doesn't consider it one of the main subsystems, a sixth system, WWKL<sub>0</sub>, will be of great use to us. It lies between RCA<sub>0</sub> and WKL<sub>0</sub> and consists of the axioms for RCA<sub>0</sub> and the weak-weak König's Lemma (*WWKL*), stating that if  $T$  is a subtree of  $2^{<\mathbb{N}}$  with no infinite path, then  $\lim_n \frac{|\{\sigma \in T \mid \text{lh}(\sigma) = n\}|}{2^n} = 0$ .

Much more could be said about the origin of each of these subsystems and their relationships with pertinent mathematical programs, such as Weyl's predicativism (ACA<sub>0</sub>) or Hilbert's finitism (WKL<sub>0</sub>), but, since Chapter 1 of [26] provides all the necessary information, there is no need to discuss it here.

All the remaining definitions in this section take place in RCA<sub>0</sub>.

The basic building block of second-order arithmetic are natural numbers. As is customary, integers and rational numbers are represented as pairs of natural numbers. Furthermore, a sufficient portion of number theory is formalizable in RCA<sub>0</sub>, most importantly, it is possible to code finite sets of natural numbers as single numbers within this system. This enables us to discuss sequences of rational numbers, and consequently real numbers.

**Definition 1.1.2** *A real number is represented as a strong Cauchy sequence of rational numbers, that is, a sequence  $\langle q_n \mid n \in \mathbb{N} \rangle$  such that  $q_n \in \mathbb{Q}$  for all  $n$ , and  $\forall m \forall n (m < n \rightarrow |q_m - q_n| \leq 2^{-m})$ .*

Two real numbers represented as  $\langle q_n \rangle$  and  $\langle q'_n \rangle$  are equal if  $\forall n (|q_n - q'_n| \leq 2^{-k+1})$  and it can be shown that all algebraic operations are well-defined. If  $x$  and  $y$  are real numbers, the statements  $x = y$ ,  $x \leq y$  are  $\Pi_1$ , while  $x \neq y$ ,  $x < y$  are  $\Sigma_1$ .

The set of real numbers is the completion of the countable dense sequence  $\mathbb{Q}$ . This definition can be generalized:

**Definition 1.1.3** *A (code for a) complete separable metric space  $\hat{A}$  is presented as a nonempty set  $A \subseteq \mathbb{N}$  together with a sequence of real numbers  $A \times A \rightarrow \mathbb{R}$  such that*

1.  $d(a, a) = 0$ ,
2.  $d(a, b) = d(b, a)$ ,
3.  $d(a, b) + d(b, c) \geq d(a, c)$ ,

for  $a, b, c \in A$ .

A point of  $\hat{A}$  is a sequence  $\langle a_n \mid n \in \mathbb{N} \rangle$  such that  $d(a_n, a_m) < 2^{-m}$  for  $m < n$ . If  $a$  and  $b$  are two points in  $\hat{A}$ , given as  $\langle a_n \rangle$  and  $b = \langle b_n \rangle$ , then  $d(a, b) = \lim_n d(a_n, b_n)$ , and two points  $a$  and  $b$  are considered equal if  $d(a, b) = 0$ , which makes  $d$  a metric. Each  $a \in A$  is identified with the point  $\langle a \mid n \in \mathbb{N} \rangle \in \hat{A}$ , and  $A$  is dense in  $\hat{A}$ .

In other words, a complete separable metric space is the completion of a countable dense subset  $A$ , and the elements of such a space are strong Cauchy sequences of elements of  $A$ . This is a typical construction which corresponds to the standard procedure of completing a metric space, and we will see later that the definitions for Banach and Hilbert spaces are rather similar to this one. Note that the set of real numbers does not formally exist as an object in second-order arithmetic and neither does any metric space  $\hat{A}$ .

The name *complete* metric space is justified, because  $\text{RCA}_0$  proves the following: if  $\langle x_n \rangle$  is a sequence of elements of  $\hat{A}$  with the property that  $d(x_n, x_m) \leq r(m)$  for  $m < n$ , where  $\lim_n r(n) = 0$ , then there exists an element  $x \in \hat{A}$  such that  $d(x, x_n) \rightarrow 0$ .

**Definition 1.1.4** A code for a (partial) continuous function  $\phi : \hat{A} \rightarrow \hat{B}$  between two complete separable metric spaces  $\hat{A}$  and  $\hat{B}$  is given as a set of quintuples  $\Phi \subseteq N \times A \times \mathbb{Q}^+ \times B \times \mathbb{Q}^+$  with the following properties (write  $(a, r)\Phi(b, s)$  to abbreviate  $\exists n ((n, a, r, b, s) \in \Phi)$ ):

1. If  $(a, r)\Phi(b, s)$  and  $(a, r)\Phi(b', s')$  then  $d(b, b') \leq s + s'$  (all neighborhoods of  $\Phi$  intersect when  $x$  is within  $r$  from  $a$ ).
2. If  $(a, r)\Phi(b, s)$  and  $(a', r') < (a, r)$  (meaning that  $d(a, a') + r' < r$ ) then  $(a', r')\Phi(b, s)$  (consistency condition).
3. If  $(a, r)\Phi(b, s)$  and  $(b, s) < (b', s')$  then  $(a, r)\Phi(b', s')$  (if  $\phi(x)$  is in an open ball, it is also in all open balls containing it).

The formula  $(a, r)\Phi(b, s)$  intuitively means that if  $x \in B(a, r)$ , then  $\phi(x)$  is in the closure of  $B(b, s)$  (provided it is defined) and all such formulas give enough information about  $\Phi(x)$ .

A point  $x \in \hat{A}$  is said to be in the domain of  $\phi$  if for all  $\epsilon$  there exists  $(a, r)\Phi(b, s)$  such that  $d(x, a) < r$  and  $s < \epsilon$ . It is provable within  $\text{RCA}_0$  that if  $x \in \text{dom}(\phi)$ , there is a unique point  $y \in \hat{B}$  such that  $d(y, b) \leq s$  for all  $(a, r)\Phi(b, s)$  and  $d(x, a) < r$ .



We will have little use for this definition. A much more important space of functions will be the spaces  $C(X)$  defined in Section 4.1 on page 42, and can be thought of as the space of all functions on a compact space with moduli of uniform continuity.

**Definition 1.1.5** *A continuous function  $f : \hat{A} \rightarrow \hat{B}$  between two metric spaces is uniformly continuous if for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that whenever  $d(x, y) < \delta$  then  $d(f(x), f(y)) < \varepsilon$ . If there exists a function  $h : \mathbb{N} \rightarrow \mathbb{N}$  such that  $d(x, y) < h(n) \rightarrow d(f(x), f(y)) < 2^{-n}$ , the function  $h$  is the modulus of uniform continuity.*

**Definition 1.1.6** *(A code for) an open subset  $U$  of a complete separable metric space is given as  $U \subseteq \mathbb{N} \times A \times \mathbb{Q}^+$  and a point  $x \in \hat{A}$  is said to belong to  $U$  if*

$$\exists n \exists a \exists r (d(x, a) < r \wedge (n, a, r) \in U).$$

The formula encodes the open set which is the (countable) union of open balls  $B(a, r)$  with rational radii, and centered at the points of  $A$ . It is provable in  $\text{RCA}_0$  that finite intersections and countable unions of open sets are open. The formula  $x \in U$ , where  $U$  is an open set, is  $\Sigma_1$ .

A set is *closed* if it is the complement of an open set. A closed set  $C$  is represented with the same code as an open set  $U$ , and  $x \in C$  iff  $x \notin U$ .

This representation of closed sets is not always useful, since it only provides negative information. Another definition will be given later (Definition 2.1.7, page 18), one that will be more convenient to work with. To prove that the two notions coincide, stronger set existence axioms,  $(ACA)$  and  $(\Pi_1^1-CA)$  are needed.

An important property that a set may have is locatedness. It will provide additional information about the set in question, as will be shown in Section 2.2. The definition applies to all sets, though for our purposes, the only relevant case is when  $C$  is closed.

**Definition 1.1.7** *A subset  $C$  of a complete separable metric space  $X$  is said to be located if there exists a continuous function  $f$  such that  $f(x) = d(x, C)$  for all  $x \in X$ . The function  $f$  is called a locating function.*

The Heine-Borel characterization of compactness and sequential compactness are both inconvenient from the point of view of second-order arithmetic, as the former is equivalent to  $(WKL)$  and the latter to  $(ACA)$  over  $\text{RCA}_0$ , which means that having one or the other would be the same as having weak König's lemma or arithmetic comprehension, respectively, available in

$\text{RCA}_0$ , which is obviously undesirable. The definition below coincides with that used by Bishop.

**Definition 1.1.8** *A compact metric space is a complete separable metric space  $\hat{A}$  such that there is an infinite sequence of finite sequences  $\langle x_{ij} \mid i \leq n_j, j \in \mathbb{N} \rangle$  of points from  $\hat{A}$  with the property that*

$$\forall z \in \hat{A} \forall j \in \mathbb{N} \exists i \leq n_j (d(x_{ij}, z) < 2^{-j}).$$

## 1.2 Some Useful Facts

The results in this section were established earlier. Proofs of the first two can be found in [2], as well as the proof of Proposition 1.2.3, except for (1)  $\rightarrow$  (2), which is given in [26]. They are simple facts with straightforward proofs that will be used throughout.

**Lemma 1.2.1** ( $\text{RCA}_0$ ) *Let  $\varphi(x, y)$  be any  $\Sigma_1$  formula, possibly with set and number parameters other than the ones shown. Then*

$$\forall x \exists y \varphi(x, y) \rightarrow \exists f \forall x \varphi(x, f(x)).$$

We will also make use of least element principles. The following claim states that in  $\text{RCA}_0$  the least element principle is available for  $\Sigma_1$  (and  $\Pi_1$ ) formulas. The class of  $\Delta_0(\Sigma_1)$  formulas is defined to be the smallest class of formulas containing the  $\Sigma_1$  formulas and closed under Boolean operations (including negation) and bounded quantification.

**Proposition 1.2.2** ( $\text{RCA}_0$ ) *The following induction principles are derivable:*

1. *Ordinary induction for  $\Delta_0(\Sigma_1)$  formulas.*
2. *Complete induction for  $\Delta_0(\Sigma_1)$  formulas.*
3. *The least-element principle for  $\Delta_0(\Sigma_1)$  formulas.*

Reversals are always proved using one of the following equivalent characterizations.

The Turing jump of  $Z \subseteq \mathbb{N}$  is defined as the set  $\{x \mid \exists y \theta(x, y, Z)\}$ , where  $\theta$  is  $\Delta_0$  and  $\exists y \theta(x, y, Z)$  is a complete  $\Sigma_1$  formula.

**Proposition 1.2.3** ( $\text{RCA}_0$ ) *The following statements are equivalent:*

1. Every increasing sequence  $\langle a_n \rangle$  of real numbers in  $[0, 1]$  has a limit.
2. If  $\langle b_n \rangle$  is any sequence of nonnegative reals such that for each  $n$ ,  $\sum_{i < n} b_i \leq 1$ , then  $\sum_n b_n$  exists.
3. If  $\langle c_n \rangle$  is any sequence of reals such that for each  $n$ ,  $\sum_{i < n} c_i^2 \leq 1$ , then  $\sum_n c_n^2$  exists.
4. If  $\langle d_n \rangle$  is any sequence of real numbers, the set  $D = \{i \in \mathbb{N} \mid d_i \neq 0\}$  exists.
5. For every  $Z \subseteq \mathbb{N}$ , the Turing jump of  $Z$  exists.
6. (ACA)

### 1.3 Brief Overview of Ergodic Theory

The original motivation for ergodic theory comes from the following problem from physics:

Consider a mechanical system with  $n$  degrees of freedom. . . . the system consists of  $k$  particles in a vessel in three-dimensional space. Assuming that the masses of these particles are completely known, the instantaneous state of the system can be described by giving the values of the  $n$  coordinates of position together with the corresponding  $n$  velocities. . . . The state of the system becomes from this point of view a point in  $2n$  dimensional space. As time goes on, the state of the system changes in accordance with the appropriate physical laws. . . . In accordance with classical, deterministic, mechanics, that entire trajectory can, in principle, be determined, once one point of it is given. In practice we almost never have enough information for such a complete determination. . . . Instead of asking “what will the state of the system be at time  $t$ ?”, we should ask “what is the probability that at time  $t$  the state of the system will belong to a specified subset of phase space?”. The questions of greatest interest are the asymptotic ones: what will (probably) happen to the system as  $t$  tends to infinity? (Halmos, [16, p.1-2])

To simplify matters, instead of considering continuous time flow in the problem above, the system is observed at discrete moments. To rephrase the last question: what do we know about the convergence of  $S_n(x) =$

$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$ , where  $f$  represents “any numerical quantity determined by the momentary state of the mechanical system (for instance, the force exerted by the given system, assumed to be a large collection of gas molecules contained in a vessel)” (Dunford and Schwartz, [12, p.657]). Based on a theorem by Liouville, “if the coordinates used in the description of the phase space are appropriately chosen, then the flow in phase space leaves all volumes invariant. In other words, the transformations that constitute the flow are measure preserving transformations...” (Halmos, [16, p.2]). It is therefore reasonable to suppose that  $\mu(T^{-1}(E)) = \mu(E)$  if  $E$  is a measurable set. If  $f$  is a characteristic function of a set  $E$ , the limit of the average above (if it exists) can be thought of as the mean amount of time the point  $x$  spends in  $E$ .

This question occupied physicists at the end of the 19<sup>th</sup> and beginning of the 20<sup>th</sup> century; as its appeal to them began to wane, it became increasingly interesting to mathematicians. Ergodic theory especially flourished between 1930 and 1960. Two of the earliest results in ergodic theory were mean and pointwise ergodic theorems, the subject of this treatise. The mean ergodic theorem was proved first, by von Neumann in 1931 [28], but he did not publish it until the following year. It established convergence in the mean of the aforementioned sequence of averages  $\langle S_n f \mid n \in \mathbb{N} \rangle$  for all square integrable functions. The pointwise ergodic theorem (also referred to as Birkhoff’s, or sometimes as Birkhoff-Khinchine’s ergodic theorem) the more “prestigious” result (von Neumann’s biographer refers to the mean ergodic theorem as the “naive” version of the pointwise) was proved and published by Birkhoff [4] in 1931. It showed almost everywhere convergence of the sequence  $\langle S_n f \rangle$  for every integrable function  $f$ . Proofs of both theorems considered standard today are due to Riesz [23].

Ergodic theory has become a prominent area of mathematics. Requirements on the systems considered have been altered and weakened and there are many modifications and variations of the theorems. Physicists are mostly interested in ergodic transformations, those in which the only invariant sets have measure 0 or 1, and much research has been done in finding necessary and sufficient conditions for ergodicity.

This dissertation is only concerned with the two main theorems in their most basic form. The sources used can be divided into two groups: classical and constructive. The relevant classical sources, apart from [16] are [3, 12, 21, 29]. The most important constructive sources for the ergodic theorems are [6, 5, 7, 20] and especially [27]. An extensive historical introduction into the subject (and the controversies connected to it) is given in [13].

## 1.4 Challenges of Reverse Mathematics

The goal of a reverse mathematician is to code as large a portion of mathematics as possible within second-order arithmetic, utilizing only natural numbers and sets of natural numbers. This endeavor begins with Gödel's sequencing function, which allows us to code sequences of natural numbers (hence integers and rationals), and consequently more complicated objects we will deal with, with natural numbers. In the construction every finite object is assigned a single natural number, its code. It is to be expected that this alters and limits discourse, even if we are only interested in countable mathematics. A consequence of this limitation is that, because quantification is permitted only over numbers and sets, it is not possible to consider equivalence classes of infinitary objects, so for example, a real number is not the equivalence class of all Cauchy sequences convergent to that number. Instead, any such sequence represents that number. Similarly, an integrable function has different representations. For our theory to be sound, it is crucial that results do not depend on representations of objects. For most considerations, this issue does not come up or it is trivial to resolve it. The only nontrivial occurrence of this concern we will come across here is when dealing with integrable functions, more specifically with their products and powers, and we will deal with it accordingly.

The greatest challenge, however, lies in formalizing familiar concepts in this restricted language. How do we choose definitions? Set theory often-times provides a number of classically equivalent definitions for a term, some of which are not compatible with the countability requirement, and even if they are, it may not be clear which one to use. How is this decided? How do we know that the definition we end up choosing is the best one? And which criteria are used to decide what "best" means?

For example, there are multiple ways to introduce the real numbers. We will opt for Cauchy sequences with a fixed rate of convergence. What would happen if another definition was chosen? Would the results obtained be different? Most of the time these questions can be answered and different approaches can be compared. In the example of compact spaces (and more generally), we want definitions with least logical strength. Choosing a different definition of compactness from the one we have would change the results associated with it. Or, consider closed sets. By choosing different definitions, we will have different outcomes regarding orthogonal projections. If the set under consideration is closed in the sense given in Section 1.1, it is not even clear if projection exists as an operator. Sometimes the question is not which definition is preferable, but whether it can be used at all. For

example, it is not possible to define an arbitrary function; and even when we restrict our attention to continuous functions, they are still not given via the usual epsilon-delta characterization, but in terms of countable representations. Though counterintuitive, this definition is consistent with the goal of assigning every object its code.

In reverse mathematics, as in the study of constructive and recursive mathematics, it is common to insist that objects come equipped with such additional information, especially when such information is typically available. In the prolog to Bishop and Bridges' *Constructive Mathematics*, Bishop refers to this practice as the "avoidance of pseudogenerality."

Of course, in many instances, the choice of formal definition is more-or-less canonical, or various natural definitions can be shown to have equivalent properties in a weak base theory... We contend, however, that in situations where there are a plurality of inequivalent "natural" representations of mathematical notions, this should not be viewed as a bad thing. Indeed, the nuances and bifurcations that arise constitute much of the subject's appeal! Set theoretic foundations provide a remarkably uniform language for communicating mathematical concepts, as well as powerful principles to aid in their analysis; but from the point of view of the mathematical logician, this uniformity and power can sometimes obscure interesting methodological issues with respect to the way the concepts are actually used. (Avigad and Simic, [2, p.3])

It is to be expected that difficulties will arise with respect to the proofs of both ergodic theorems. Here are some of the problems:

*(Mean ergodic theorem)* Classically, each Hilbert space can be decomposed into subspaces  $M$  and  $N$  (defined later) which are orthogonal to each other. The limit of the sequence of operators  $\langle S_n \rangle$  is the projection operator on  $M$  (denoted as  $P_M$ ) and (classically) the existence of  $P_M$  implies the existence of the limit. Showing that  $N$  is the orthogonal complement of  $M$  is not trivial (or, generally, provable in  $\text{RCA}_0$ ), and the existence of  $P_M$  may not imply convergence of  $\langle S_n \rangle$  at all.

*(Pointwise ergodic theorem)* Classically, the measure preserving transformation  $T$  is defined on the points of the space  $X$ . For a number

of reasons, we are unable to do the same. Instead, we consider the transformation that  $T$  induces on the space  $L_1(X)$ .

*(Pointwise ergodic theorem)* It may not be clear which properties the induced transformation need have. Classically, it is multiplicative, non-negative, preserving  $L_1$  and  $L_\infty$  norm. With our definition it will turn out that some of these properties imply others, and that the assumptions made also depend on the proof.

*(Pointwise ergodic theorem)* The standard proof of the pointwise ergodic theorem is more difficult to formalize than the other two. This proof, first given by Riesz, is classically short and simple. However, it requires the knowledge that a certain set is invariant, which is the main source of difficulties for us, because second-order arithmetic for the most part takes a point-free approach to measure. Talking about invariant sets in our framework is difficult, and showing that a set (or function) is invariant requires using methods borrowed from the classical theory of integrable functions, so the proof ends up being less natural than the other two.

These issues will be discussed in more detail in the appropriate chapters.

## 1.5 Parallels With Constructive Mathematics

There is a close connection between mathematics formalized in subsystems of second-order arithmetic and constructive mathematics, especially between  $\text{RCA}_0$  and Bishop-style mathematics. Though the similarity between  $\text{RCA}_0$  and computable mathematics is probably even more pronounced, we consider the work of Bishop and Bridges, as it is more complete and covers all topics we will discuss. Furthermore, a number of their proofs have been invaluable in my work.

Simpson pointed out the main differences between the two approaches in [26, p.31]. Apart from the immediate difference that constructivists presuppose intuitionist logic (which is why, for example the intermediate value theorem fails constructively, but is provable in  $\text{RCA}_0$ ), the main distinction is that Bishop never specified a formal system to work in. There are no restrictions in the complexity of formulas used, induction is not restricted, and sets are defined somewhat vaguely. According to Bishop [6, p.7],

A set is defined by describing exactly what must be done in order to construct an element of the set and what must be done

in order to show that two elements are equal... There is always ambiguity, but it becomes less and less as the reader continues to read and discover more and more of the author's intent, modifying his interpretations if necessary to fit the intentions of the author as they continue to unfold.

Bridges [9, p.5] states "Thus, if  $P(x)$  is some property of  $x$ , we can form the class  $\{x \in A \mid P(x)\}$  of those  $x$  in  $A$  for which  $P(x)$  holds." He then concludes "Some people find it hard to accept our freedom of construction of subclasses by abstraction." In other words, constructivists allow themselves freedom in definitions of sets that we do not have. The maximal ergodic theorem is an example of a theorem that is constructively valid, but needs to be proved in  $ACA_0$  in our framework. Similarly, Bishop defines closures of sets without hesitation. For us, even the statement that an open set in a general metric space has a closure is equivalent to  $(II_1^1 - CA_0)$ .

The other cause for discrepancies comes from difference in definitions. To state just one example, a uniformly continuous function in [8] always comes equipped with a modulus of uniform continuity, whereas we need  $WKL_0$  to show that a uniformly continuous function has a modulus. More often than not, though, definitions in the two frameworks coincide, like that for a real number or a compact space, and most constructive proofs are easily translated, for example existence of orthogonal projection or the Riesz representation theorem. In fact, encountering a result of Bishop and Bridges which cannot be translated directly into  $RCA_0$  is an anomaly. For this reason, the question of differences between the two frameworks is more interesting than that of similarities.

Either way, the exploration of these issues is a worthy task, but one that would require much time and space. We will instead only look into some questions concerning measure theory and the pointwise ergodic theorem, and we will do so only in passing. An interesting situation occurs with respect to measure theory. Bishop and Bridges call a set that contains the domain of an integrable function *full* and are able to show that the complement of a full set has measure 0, which is quite similar to the statement that an integrable function is defined almost everywhere. In comparison, Simpson and Yu showed (see [30, 11]) that to show this fact in our framework, one needs to presuppose  $(WWKL)$ . More precisely,  $RCA_0$  proves that an integrable function is defined outside of a  $G_\delta$  set, while with  $(WWKL)$  it can be shown that this set is null. Where does this difference stem from?

Both approaches to integrability are based on the classical Daniell integral. In particular, Bishop lists axioms that every *integration space* need



satisfy and proves that  $C(X)$  (the space used to build integrable functions) satisfies them. Once this fact is established, it is easy to show that a complement of a full set has measure 0. What about our framework? It cannot be shown in  $\text{RCA}_0$  that  $C(X)$  satisfies the classical definition of an integration space and Appendix A will contain a proof of this fact, but it is still unclear whether the proof by Bishop and Bridges that  $C(X)$  satisfies the constructive definition can be formalized. It is likely that it cannot be used directly, as it relies on properties of sets that we do not have. The other discrepancy is in the definition of complements: they replace the notion of the complement of a set with that of *complemented sets*, pairs of sets that do not intersect. They do not prove that the domain of an integrable function is complemented, meaning that it may not always have a characteristic function. Before giving any definite answers, however, a much more careful analysis of the two approaches needs to be conducted.

The ergodic theorems are not provable constructively. In [6], Bishop gives an example of two equal tanks, with  $T$  representing the motion of the fluid inside them. Depending on whether there is a leak between the tanks or not, the system will exhibit perceptibly different behavior after a sufficient amount of time. We do not know beforehand if there is a leak or not, so different limits of the sequence  $\langle S_n \rangle$  as defined in the previous section imply the law of the excluded middle (we will see later that even assuming the law of the excluded middle does not save us, and that proofs of both the mean and pointwise ergodic theorem require arithmetic comprehension). Bishop gave three different proofs (all equally complicated) of the pointwise ergodic theorem. Each used *upcrossing inequalities* and provided what is called an *equal hypothesis* substitute for the pointwise ergodic theorem, meaning that the conclusion is classically equivalent to the conclusion of the ergodic theorem, but is constructively weaker. Nuber [20], his student, provided a proof which obtained the classical result, but with stronger assumptions. The proof relied heavily on functional analysis, and was almost equally complicated. Finally, Spitters [27] provided a comprehensible constructive version of the ergodic theorems. His proof of the mean ergodic theorem is similar to the standard one and in a number of cases his results provided a foundation for my proofs.



## Chapter 2

# Banach and Hilbert Spaces

### 2.1 Preliminaries

Many of the relevant definitions were given in the introduction. Concepts introduced below pertain to Banach and Hilbert spaces that will be used in this and the next chapter. As before, all definitions are made in  $\text{RCA}_0$ .

**Definition 2.1.1** *A countable vector space  $A$  over a countable field  $K$  consists of a set  $|A| \subseteq \mathbb{N}$  with operations  $+: |A| \times |A| \rightarrow |A|$  and  $\cdot: |K| \times |A| \rightarrow |A|$ , and a distinguished element  $0 \in |A|$  such that  $(|A|, +, \cdot)$  satisfies the usual properties of a vector space over  $K$ .*

**Definition 2.1.2** *A (code for a) separable Banach space  $\hat{A}$  consists of a countable vector space  $A$  over  $\mathbb{Q}$  together with a sequence of real numbers  $\|\cdot\|: A \rightarrow \mathbb{R}$  such that*

1.  $\|q \cdot a\| = |q| \|a\|$  for all  $q \in \mathbb{Q}$  and  $a \in A$ ,
2.  $\|a + b\| \leq \|a\| + \|b\|$  for all  $a, b \in A$ .

The construction is similar to that of a complete separable metric space. A point of  $\hat{A}$  is a sequence  $\langle a_n \mid n \in \mathbb{N} \rangle$  of points in  $A$  such that  $\|a_m - a_n\| \leq 2^{-m}$  for all  $m < n$ . The function  $d(a, b) = \|a - b\|$  is a pseudometric on  $A$ . The norm and metric can be extended to all of  $\hat{A}$  and the resulting space is a complete separable normed space.

**Definition 2.1.3** *A code for a bounded linear operator between separable Banach spaces  $\hat{A}$  and  $\hat{B}$ , is a sequence  $F: A \rightarrow \hat{B}$  of points of  $\hat{B}$  indexed with elements of  $A$ , such that*

1.  $F(q_1 a_1 + q_2 a_2) = q_1 F(a_1) + q_2 F(a_2)$  for all  $q_1, q_2 \in \mathbb{Q}$  and  $a_1, a_2 \in A$ ,
2. There exists a real number  $\alpha$  such that  $\|F(a)\| \leq \alpha \|a\|$  for all  $a \in A$ .

Then, for  $x$  given as  $\langle a_n \in \mathbb{N} \rangle$  define  $F(x) = \lim_n F(a_n)$  (limit exists in  $\text{RCA}_0$  by virtue of boundedness of  $F$ ) and for  $x \in \hat{A}$ ,  $\|F(x)\| \leq \alpha \|x\|$ .

If continuous linear operators are defined as total continuous functions which are further linear,  $\text{RCA}_0$  proves that the two notions coincide: a linear operator is bounded if and only if it is continuous.

The definition of a Hilbert space differs little from that of a Banach space.

**Definition 2.1.4** A real separable Hilbert space  $H$  consists of a countable vector space  $A$  over  $\mathbb{Q}$  together with a function  $\langle \cdot, \cdot \rangle : A \times A \rightarrow \mathbb{R}$  satisfying

1.  $\langle x, x \rangle \geq 0$ ,
2.  $\langle x, y \rangle = \langle y, x \rangle$ ,
3.  $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$ ,

for all  $x, y, z \in A$  and  $a, b \in \mathbb{Q}$ .

Given a Hilbert space  $H$  as above, define a function  $d(x, y)$  on  $A$  by  $d(x, y) = \langle x - y, x - y \rangle^{1/2}$ . As before,  $d$  is a pseudometric and  $H$  is the completion of  $A$  under this pseudometric, which is often written as  $H = \hat{A}$ . The inner product is extended to the whole space by letting  $\langle x, y \rangle = \lim_n \langle x_n, y_n \rangle$  for  $x$  and  $y$  represented as  $\langle x_n \rangle$  and  $\langle y_n \rangle$  respectively. The inner product can be shown to be continuous. Once again, define  $x = y$  to mean that  $d(x, y) = 0$ , making  $d$  a metric. The space  $H$  is separable and Cauchy complete.

The norm function on a Hilbert space is defined by  $\|x\| = \langle x, x \rangle^{1/2} = d(x, 0)$ , so every Hilbert space gives rise to a Banach space, in the sense of reverse mathematics.

Clearly, every statement that holds for all Banach spaces will hold for Hilbert spaces as well. Because Hilbert spaces have a richer geometric structure, however, the reverse is not true. There exist plenty of results that are valid specifically for Hilbert spaces.

To prove the reversal of the mean ergodic theorem, complex Hilbert spaces need to be introduced. Although we haven't explicitly defined what a complex number is, it should be clear that it can be defined as an ordered pair of reals, with operations defined in the standard way. Just as the

real numbers are the completion of the rationals, complex numbers are a completion of complex rationals, the countable set  $\mathbb{Q}(i)$ , which can also be thought of as the set of all ordered pairs of rational numbers. The definition of a complex Hilbert space is as then follows:

**Definition 2.1.5** *A complex separable Hilbert space  $H$  consists of a countable vector space  $A$  over  $\mathbb{Q}(i)$  together with a function  $\langle \cdot, \cdot \rangle : A \times A \rightarrow \mathbb{R}$  satisfying*

1.  $\langle x, x \rangle \geq 0$ ,
2.  $\langle x, y \rangle = \overline{\langle y, x \rangle}$ ,
3.  $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ ,

for all  $x, y, z \in A$  and  $a, b \in \mathbb{Q}(i)$ .

The notion of a bounded linear operator can be lifted accordingly.

The standard examples of separable infinite dimensional Hilbert and Banach spaces can be developed in  $\text{RCA}_0$ . Examples of Hilbert spaces include the space  $L_2(X)$  of square-integrable real-valued functions on any compact separable metric space  $X$ , and the space  $l_2$  of square-summable sequences of reals. An example of a complex Hilbert space is the space of square-summable sequences of complex numbers  $l_2(\mathbb{C})$ . Standard examples of Banach spaces are the space of uniformly continuous functions with a modulus of uniform continuity  $C(X)$ , and  $L_1(X)$ , the space of integrable real-valued functions over  $X$ . All of these spaces will occur in later chapters.

The mean ergodic theorem is usually stated for a transformation on a Hilbert space which is an isometry, but it is enough that it be nonexpansive. An isometry is a bounded linear operator that preserves norm, so the definition is analogous to that of a bounded linear operator.

**Definition 2.1.6** *An isometry on a Banach (Hilbert) space  $H = \hat{A}$  is a mapping  $T : A \rightarrow H$  such that*

1.  $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$  for  $\alpha, \beta \in \mathbb{Q}$  and  $x, y \in A$ ,
2.  $\|Tx\| = \|x\|$  for all  $x \in A$ .

A transformation is nonexpansive if  $\|Tx\| \leq \|x\|$  holds for all  $x \in A$  instead of clause 2. above.

Since we will be looking at iterations of the transformation  $T$ , we will need to know that  $T^n$  is also an isometry (or nonexpansive) for all  $n$ . Using

Lemma 1.2.1 we can make sense of the sequence of iterations  $T^n$ , and hence of the partial averages  $\langle S_n \rangle$ . For every  $x$  and  $y$  the following holds:

$$\|Tx - Ty\| = \|T(x - y)\| \leq \|x - y\|.$$

Since  $T$  is a continuous function, the identity is a modulus of continuity for  $T$ . It is shown in [14] that if a function has a modulus of uniform continuity, then  $T^n$  is continuous for all  $n$ . Since the statement that  $T^n$  is linear (resp. an isometry, nonexpansive operator) is  $\Pi_1$ , and since  $\Pi_1$  induction is equivalent to  $\Sigma_1$  induction, it can be proved by induction in  $\text{RCA}_0$  that if  $T$  is linear (resp. an isometry, nonexpansive linear operator) then so is  $T^n$  for each  $n$ .

Let us now return to closed sets. As was already mentioned, there is more than one notion of closed to consider. A more useful concept from a closed set being the complement of an open is that of a separably closed set.

**Definition 2.1.7** *The code for a separably closed subset  $S$  of a metric space  $X = \hat{A}$  is given as a countable sequence  $\langle x_n \mid n \in \mathbb{N} \rangle$  of points in  $A$ .*

The set  $S$  is the completion of the sequence  $\langle x_n \rangle$ , therefore  $x \in S$  means  $\forall m \exists n (d(x, x_n) < 2^{-m})$ .

**Definition 2.1.8** *A closed subspace of a Hilbert (resp. Banach) space is a subset of the space which is a Hilbert (resp. Banach) space in itself.*

*Note:* Upon inspection, it can be seen that a separably closed subset of a Banach or Hilbert space is a closed subspace of the space and conversely, i.e. the two definitions are equivalent and can be used interchangeably.

For orthogonal projection onto a set to exist, the set in question needs to be linear. Consider the following four notions for a Banach space  $X$ .

1. A *closed linear set* is a closed subset of  $X$  that is further linear.
2. A *closed subspace* is a subset of  $X$  which is a Banach space in itself.
3. A *located closed linear set* is a closed linear set with a locating function.
4. A *located closed subspace* is a closed subspace with a locating function.

These definitions are not vacuous, nor are the relationships among them trivial. We showed in [2] that the third and fourth notion in the above definition coincide in  $\text{RCA}_0$ . All others are distinguishable in  $\text{RCA}_0$  and occur in practice. For example, if  $f$  is a bounded linear functional, its kernel,  $\ker f$ , is a closed linear set; it is provable in  $\text{RCA}_0$  that it is also a subspace,

and that it is located if and only if the norm of  $f$  exists. In the proof of the mean ergodic theorem we will consider the set  $\{x \mid Tx = x\}$ . This set is closed and linear. However, for an appropriately chosen  $T$ , (ACA) will be required to show that it is a closed subspace. Similarly, the set spanned by  $\{x - Tx \mid x \in A\}$  is a closed subspace, but, as above, (ACA) may be necessary to prove it is a closed set.

**Definition 2.1.9** *Let  $M$  be a closed subspace of a Hilbert space  $H$ . Let  $x$  and  $y$  be elements of  $H$ , with  $y$  in  $M$ . If  $d(x, y) \leq d(x, z)$  for any other point  $z$  in  $M$ ,  $y$  is said to be the projection of  $x$  on  $M$ . Let  $P$  be a bounded linear operator from  $H$  to itself. If for every  $x$  in  $H$ ,  $Px$  is the projection of  $x$  on  $M$ ,  $P$  is said to be the projection function for  $M$ .*

Although existence of orthogonal projections is typically associated with Hilbert spaces, there are certain Banach spaces that have this property, for example, uniformly convex spaces. Classically, a Banach space is uniformly convex if it has the additional property that for every two sequences of vectors  $\langle x_n \rangle$  and  $\langle y_n \rangle$ , if  $\|x_n\| \rightarrow 1$ ,  $\|y_n\| \rightarrow 1$  and  $\|x_n + y_n\| \rightarrow 2$ , then  $\|x_n - y_n\| \rightarrow 0$ . For our purposes, a *modulus of convexity* is needed, giving the rate of convergence of the norms. The modified definition is:

**Definition 2.1.10** *A Banach space is uniformly convex if there exists a function  $\delta : \mathbb{Q} \rightarrow \mathbb{Q}$  such that whenever  $\|u\| \leq 1$ ,  $\|v\| \leq 1$  and  $\|u + v\| \geq 2 - \delta(\varepsilon)$ , then  $\|u - v\| \leq \varepsilon$ .*

This property is a substitute for the parallelogram identity, which does not hold in Banach spaces without an inner product.

## 2.2 Properties of Banach and Hilbert Spaces

In [2] we showed that every Hilbert space has an orthonormal basis in  $\text{RCA}_0$ . Since it has no bearing on the proof of the mean ergodic theorem, this proof is omitted.

Let us now establish the relationship between the notions of “closed” and “separably closed”. In the case of metric spaces, the work has been done by Brown [10].

**Theorem 2.2.1** ( $\text{RCA}_0$ ) *Each of the following statements is equivalent to (ACA):*

1. *In a compact metric space, every closed set is separably closed.*

2. In  $[0, 1]$ , every closed set is separably closed.
3. In an arbitrary separable metric space, every separably closed set is closed.
4. In  $[0, 1]$ , every separably closed set is closed.

Each of the following statements is equivalent to  $(\Pi_1^1\text{-CA})$ :

1. In an arbitrary space, every closed set is separably closed.
2. In Baire space  $(\mathbb{N}^{\mathbb{N}})$ , every closed set is separably closed.

The following theorem collects some of the results obtained in [2]. In a sense, it is the analogue of Theorem 2.2.1 for Banach (and Hilbert) spaces.

**Theorem 2.2.2** ( $\text{RCA}_0$ ) *With respect to Banach spaces, the following hold:*

1. *The statement that every closed subspace is a closed linear set is equivalent to (ACA).*
2. *The statement that every closed linear set is a closed subspace is implied by  $(\Pi_1^1\text{-CA})$  and implies (ACA).*
3. *Every located closed subspace is a located closed linear set and conversely.*
4. *The statement that every closed subspace is located is equivalent to (ACA).*
5. *The statement that every closed linear set is located is implied by  $(\Pi_1^1\text{-CA})$  and implies (ACA).*

Notice that the exact strength of 2. and 5. is not known. This question, along with a number of others, remains open. See also Chapter 6 or Section 16 of [2].

The above theorem also shows that proving that a certain set is located usually requires strong set existence axioms. The following results hold for metric spaces and are available in [10].

**Theorem 2.2.3** ( $\text{RCA}_0$ ) *Each of the following statements is equivalent to (ACA):*

1. *Every nonempty closed set in a compact space is located.*
2. *Every nonempty closed set in  $[0, 1]$  is located.*



3. Every nonempty separably closed set in an arbitrary space is located.
4. Every nonempty separably closed set in  $[0, 1]$  is located.
5. Every nonempty open set in an arbitrary space is located.
6. Every nonempty open set in  $[0, 1]$  is located.

Each of the following is equivalent to  $(\Pi_1^1\text{-CA})$ :

1. Every nonempty closed set in an arbitrary space is located.
2. Every nonempty closed set in Baire space is located.

*Note:* It was mentioned earlier that the code for a closed subspace provides more information than the code for a closed set. The previous theorem justifies that claim, as proving that a separably closed set is located requires weaker axioms than proving the same for a closed set ( $(ACA)$  vs.  $(\Pi_1^1\text{-CA})$ ).

In [2]) we were able to provide an improvement to the above theorem, showing that even existence of distances for individual points requires strong axioms; more precisely, all relationships remain the same when the requirement that a set is located is replaced with existence of distance from any point to that set.

**Theorem 2.2.4** ( $\text{RCA}_0$ ) *Each of the following is equivalent to  $(ACA)$ :*

1. In a compact space, if  $C$  is any nonempty closed set and  $x$  is any point, then  $d(x, C)$  exists.
2. If  $C$  is any nonempty closed subset of  $[0, 1]$ , then  $d(0, C)$  exists.
3. In an arbitrary space, if  $O$  is a nonempty open set and  $x$  is any point, then  $d(x, O)$  exists.
4. If  $O$  is any nonempty subset of  $[0, 1]$ , then  $d(0, O)$  exists.

Each of the following is equivalent to  $(\Pi_1^1\text{-CA})$ :

1. In an arbitrary space, if  $C$  is any closed set and  $x$  is any point, then  $d(x, C)$  exists.
2. In any compact space, if  $S$  is any  $G_\delta$  set and  $x$  is any point, then  $d(x, S)$  exists.
3. If  $S$  is a  $G_\delta$  subset of  $[0, 1]$ , then  $d(0, S)$  exists.

We now restrict our attention to Hilbert spaces and uniformly convex Banach spaces and look into orthogonal projections. The notion of projection makes sense for both separably closed and closed linear sets. We will attempt to determine the strength of the statements “orthogonal projection of  $x$  onto  $M$  exists” and “orthogonal projection onto  $M$  exists” in both cases.

**Proposition 2.2.5** (RCA<sub>0</sub>) *If the projection of  $x$  on  $M$  exists, it is unique (up to equality in the Hilbert or Banach space).*

*Proof.* Note that if  $y$  is the projection of  $x$  on  $M$ , then  $\|x - y\| = d(x, M)$ . The claim is proved using the parallelogram identity

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Suppose  $y$  and  $y'$  are both projections of  $x$  on  $M$ . Then  $\|x - y\| = \|x - y'\|$ , and by linearity  $\frac{1}{2}(y + y')$  is also in  $M$ . But then

$$\begin{aligned} \|x - y\| &\leq \|x - \frac{1}{2}(y + y')\| = \|\frac{1}{2}(x - y) + \frac{1}{2}(x - y')\| \\ &\leq \frac{1}{2}\|x - y\| + \frac{1}{2}\|x - y'\| \\ &= \|x - y\|, \end{aligned}$$

hence all inequalities are actually equalities. Let  $d = \|x - y\|$ . The previous paragraph and the parallelogram identity imply

$$\begin{aligned} 4d^2 &= \|(x - y) + (x - y')\|^2 \\ &= 2(\|x - y\|^2 + \|x - y'\|^2) - \|y - y'\|^2 \\ &= 4d^2 - \|y - y'\|^2, \end{aligned}$$

so  $\|y - y'\| = 0$  and  $y = y'$ . □

The following criterion will be useful later:

**Lemma 2.2.6** (RCA<sub>0</sub>) *An element  $y$  is the projection of  $x$  on  $M$  if and only if  $y$  is in  $M$ ,  $x - y$  is orthogonal to  $M$  and  $y$  is the unique vector with these properties.*

*Proof.* If  $y$  is the projection of  $x$  onto  $M$ , then  $x - y$  is orthogonal to  $M$ . For, suppose that  $y = Px$  and  $x - y$  is not orthogonal to  $M$ . Let  $z \in M$ ,

such that  $\langle x - y, z \rangle \neq 0$ . Then for every  $a \in \mathbb{R}$ ,  $y - az$  is also an element of  $M$ , and

$$\langle x - y + az, x - y + az \rangle \geq d(x, M)^2 = \langle x - y, x - y \rangle,$$

therefore

$$|a|^2 \|z\|^2 + 2a \langle x - y, z \rangle \geq 0.$$

By taking  $a$  with sufficiently small absolute value the last inequality can be contradicted, so  $\langle x - y, z \rangle = 0$ .

To show there can be at most one point  $y$  such that  $x - y \perp M$ , assume for the sake of contradiction that there is a  $y'$  with the same property. Then

$$\begin{aligned} \langle x - y, y \rangle = 0 &\rightarrow \langle x, y \rangle = \langle y, y \rangle, \\ \langle x - y', y' \rangle = 0 &\rightarrow \langle x, y' \rangle = \langle y', y' \rangle, \\ \langle x - y, y' \rangle = 0 &\rightarrow \langle x, y' \rangle = \langle y, y' \rangle, \\ \langle x - y', y \rangle = 0 &\rightarrow \langle x, y \rangle = \langle y, y' \rangle. \end{aligned}$$

Therefore,  $\langle x, y \rangle = \langle x, y' \rangle = \langle y, y \rangle = \langle y, y' \rangle = \langle y', y' \rangle$  and

$$\langle y - y', y - y' \rangle = \langle y, y \rangle - 2\langle y, y' \rangle + \langle y', y' \rangle = 0,$$

so  $y = y'$ .

Conversely, if  $y$  is any vector in  $M$  with  $\langle x - y, z \rangle = 0$  for all  $z \in M$ , for any other vector  $y' \in M$ ,

$$\|x - y'\|^2 = \|x - y\|^2 + \|y - y'\|^2 > \|x - y\|^2,$$

and  $y$  is the closest point in  $M$  to  $x$ ; hence  $y = Px$ .  $\square$

We will use the notation  $P_M$  to denote the projection function for  $M$ , so, for example, the statement “ $P_M$  exists” is shorthand for the statement “there exists a projection function for  $M$ .” The next theorem demonstrates the relationship between distances and projections in the context of a subspace.

**Theorem 2.2.7 (RCA<sub>0</sub>)** *Let  $H$  be a Hilbert space, let  $M$  be a closed subspace of  $H$ , and let  $x$  be any element of  $H$ . Then the following are equivalent:*

1. *The distance from  $x$  to  $M$  exists.*
2. *The projection from  $x$  to  $M$  exists.*

Moreover, for any closed subspace  $M$ , the following are equivalent:

1.  *$M$  is located.*

2. The projection function,  $P_M$ , exists.

*Proof.* In both cases, the direction (2)  $\rightarrow$  (1) is immediate, since if  $y$  is the projection of  $x$  on  $M$ , then  $d(x, M) = d(x, y)$ .

For the first implication (1)  $\rightarrow$  (2), suppose  $M$  is a closed subspace with  $\langle w_m \rangle$  a dense sequence of points in  $M$ . By assumption,  $d(x, M) = \inf\{d(x, y) \mid y \in M\}$  exists.

The definition of infimum implies that

$$\forall n \exists m (d(x, w_m) < d(x, M) + 2^{-n}).$$

By Lemma 1.2.1 there is a sequence of points  $y_n$  from  $\langle w_m \rangle$  such that for every  $n$ ,

$$d(x, M) \leq d(x, y_n) < d(x, M) + 2^{-n}.$$

It is enough to show that  $\langle y_n \rangle$  is a Cauchy sequence, because this implies that  $d(x, \lim_n y_n) = d(x, M)$ . Using the parallelogram identity,

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2,$$

we have

$$\begin{aligned} \|y_n - y_m\|^2 &= \|y_n - x - (y_m - x)\|^2 \\ &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|\frac{1}{2}(y_n + y_m) - x\|^2 \\ &\leq 2(d(x, M) + 2^{-n})^2 + 2(d(x, M) + 2^{-m})^2 - 4d^2 \\ &= (2^{-n+2} + 2^{-m+2})d(x, M) + 2^{-n-m}. \end{aligned}$$

The inequality follows from the fact that  $\frac{1}{2}(y_n + y_m) \in M$  and  $d(\frac{1}{2}(y_n + y_m), x) \geq d(x, M)$ . A more careful algebraic manipulation of the last quantity would show that, given that  $m < n$ ,  $(2^{-n+2} + 2^{-m+2})d(x, M) + 2^{-n-m} \leq r(m)$ , and  $r(m) \rightarrow 0$  when  $m \rightarrow \infty$ . In  $\text{RCA}_0$  this is enough for  $\langle y_n \rangle$  to be a convergent sequence.

The second implication (1)  $\rightarrow$  (2) is just a uniform version of the preceding argument. To define the code of  $P$  as a bounded linear operator define its value at each element of the countable dense subset  $A$  of  $H$ , as above. It remains to check that  $P$  is linear and bounded. For linearity, given  $a, b \in \mathbb{Q}$  and  $x, y \in A$ ,

$$\langle ax + by - (aPx + bPy), z \rangle = a\langle x - Px, z \rangle + b\langle y - Py, z \rangle = 0,$$

so by Lemma 2.2.6

$$aPx + bPy = P(ax + by),$$

and  $P$  is linear. For any  $x \in H$ ,

$$\begin{aligned}\|x\|^2 &= \langle x, x \rangle = \langle x - Px + Px, x - Px + Px \rangle \\ &= \|x - Px\|^2 + \|Px\|^2,\end{aligned}$$

and  $\|Px\| \leq \|x\|$ . Therefore,  $\|P\| \leq 1$ .  $\square$

Note that if there is a nonzero element  $y$  in  $M$ , then  $P_y = y$  so  $\|P\| = 1$ .

If in Theorem 2.2.7 “closed subspace” is replaced by “closed linear subset,” the situation changes. The uniform case stays the same: the existence of a locating function is equivalent to the existence of a projection function.

**Theorem 2.2.8 (RCA<sub>0</sub>)** *Suppose  $M$  is a closed linear subset in a Hilbert space and  $x$  is a point in the space.*

1. *If  $P_M x$  exists, then so does  $d(x, M)$ .*
2.  *$M$  is located if and only if  $P_M$  exists.*

*Proof.* As before, it is easy to see that if  $P_M x$  exists, then  $d(x, M) = \|x - P_M x\|$ , and similarly, if  $P_M$  exists then  $f(x) = \|x - P_M x\|$  is a locating function.

It remains to show that if  $M$  is located, the projection  $y$  of  $x$  on  $M$  can be obtained uniformly. The construction is similar to the preceding one, but instead of choosing  $y_m$  in a dense subset of  $M$ , we choose it from the dense subset  $A$  of  $H$ , ensuring that  $y_m$  is close enough to  $M$ .

Let  $d$  be the locating function. By definition, for every  $n \in \mathbb{N}$  there is  $x_n \in M$  such that

$$d \leq d(x, x_n) \leq d + 2^{-n-1}.$$

By density of  $A$  in  $H$  there is  $y' \in A$  with

$$d(x_n, y') \leq 2^{-n-1},$$

which means that  $y'$  is such that

$$d - 2^{-n} \leq d(x, y') \leq d + 2^{-n}.$$

Thus we have

$$\forall n \exists y' (d - 2^{-n} < \|y' - x\| < d + 2^{-n} \wedge d(y', M) < 2^{-n}). \quad (2.1)$$

By Lemma 1.2.1 there is a sequence  $\langle y_n \rangle$  such that for each fixed  $n$ ,  $y_n$  satisfies (2.1).

The proof that  $\langle y_n \rangle$  is a Cauchy sequence is similar to that in the case of a closed subspace.

$$\begin{aligned} \|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\left\|\frac{1}{2}(y_n + y_m) - x\right\|^2 \\ &\leq (d + 2^{-n})^2 + (d + 2^{-m})^2 - 4\left\|\frac{1}{2}(y_n + y_m) - x\right\|^2. \end{aligned}$$

The only difference is in finding the lower bound for  $\left\|\frac{1}{2}(y_n + y_m) - x\right\|$ , as  $y_n$  and  $y_m$  are not necessarily in  $M$ .

Since  $d(y_n, M) < 2^{-n}$ , there is a  $w' \in M$  such that  $\|y_n - w'\| < 2^{-(n-1)}$ , and similarly there is a  $w'' \in M$  such that  $\|y_m - w''\| < 2^{-(m-1)}$ . Then  $\frac{1}{2}(w' + w'') \in M$ , so  $\left\|\frac{1}{2}(w' + w'') - x\right\| \geq d$ . Thus we have

$$\begin{aligned} \left\|\frac{1}{2}(y_n + y_m) - x\right\| &\geq \left\|\frac{1}{2}(w' + w'') - x\right\| - \left\|\frac{1}{2}(y_n + y_m) - \frac{1}{2}(w' + w'')\right\| \\ &\geq d - \frac{1}{2}\|y_n - w'\| - \frac{1}{2}\|y_m - w''\| \\ &\geq d - 2^{-n} - 2^{-m}. \end{aligned}$$

The rest of the proof remains the same:  $\langle y_n \rangle$  is a Cauchy sequence that converges to a point  $y \in H$ . Since for each  $n$ ,  $d(y_n, M) \leq 2^{-n}$ , this means  $d(y, M) = 0$ , and since  $M$  is closed, this implies  $y \in M$ . Furthermore, this  $y$  is the projection of  $x$  onto  $M$ .  $\square$

The strength of the statement “if  $d(x, M)$  exists then so does  $P_M x$ ” remains open. It is implied by  $(\Pi_I^1\text{-}CA)$ : this follows from part 2 of the above theorem, since, by Theorem 2.2.3,  $(\Pi_I^1\text{-}CA)$  suffices to prove the existence of a locating function. Whether  $\Pi_1^1$  comprehension is also necessary is not known.

We can use projections to give an alternative proof of one of the claims made in Theorem 2.2.2 above:

**Proposition 2.2.9** (RCA<sub>0</sub>) *Suppose  $M$  is closed linear set in a Hilbert space, and either  $M$  is located or  $P_M$  exists. Then  $M$  is a closed subspace.*

*Proof.* By Theorem 2.2.8, saying  $M$  is located is equivalent to saying that  $P_M$  exists. If  $A = \langle x_n \rangle$ , then  $\langle P_M x_n \rangle$  is a countable dense subset of  $M$ .  $\square$

Now that the relationship between distances and projections is known, we can ask the question of what it takes to show the existence of either. For closed subspaces of Hilbert spaces, it requires exactly  $(ACA)$ :

**Theorem 2.2.10** (RCA<sub>0</sub>) *The following are equivalent:*

1. *Every closed subspace of a Hilbert space is located.*
2. *For every closed subspace  $M$ , the projection on  $M$  exists.*
3. *For every closed subspace  $M$  and every point  $x$ , the projection of  $x$  on  $M$  exists.*
4. *For every closed subspace  $M$  and every point  $x$ ,  $d(x, M)$  exists.*
5. *(ACA).*

*Proof.* By Theorem 2.2.7 1 and 2 are equivalent, as are 3 and 4, 2 clearly implies 3 and by Theorem 2.2.3, 5 implies 1. To close the chain it suffices to show that either 2 or 4 implies 5; this will be a consequence of Theorem 3.1.3 (page 36).  $\square$

If we replace “closed subspace” by “closed linear subset,” the answer is less precise:

**Theorem 2.2.11** (RCA<sub>0</sub>) *Each of the following statements is implied by  $(\Pi_1^1\text{-CA})$  and implies (ACA):*

1. *For every closed linear subset  $M$  of a Hilbert space, the projection on  $M$  exists.*
2. *Every closed linear subset of a Hilbert space is located.*
3. *For every closed linear subset  $M$  and every point  $x$ , the projection of  $x$  on  $M$  exists.*
4. *For every closed linear subset  $M$  and every point  $x$ ,  $d(x, M)$  exists.*

*Proof.* Again, we have seen that 1 and 2 are equivalent, and it follows from Theorem 2.2.3 that they are implied by  $(\Pi_1^1\text{-CA})$ . Also, each of 1 and 2 implies 3, which in turn implies 4. The fact that 4 implies (ACA) is given by Corollary 3.2.2 below (page 39).  $\square$

What can we say about Banach spaces? If  $B$  is a uniformly convex Banach space, one can still consider the projection of  $x$  on  $M$ . It will be the closest point to  $x$  in  $M$ . The non-uniform part of Theorem 2.2.7 holds more generally for uniformly convex Banach spaces. The uniform convexity is needed in showing that  $\langle y_n \rangle$  is a Cauchy sequence, as the parallelogram

identity does not hold in Banach spaces. We cannot, however, construct a bounded linear projection operator on a Banach space. The construction in Theorem 2.2.7 cannot be paralleled: defining  $P_M x$  as the closest point to  $x$  in  $M$  when  $x$  in the dense countable subset of  $B$  does not work because to show linearity of  $P$ , an inner product is necessary — it cannot be replaced with uniform convexity.

### 2.3 Bounded Linear Functionals

A bounded linear functional on a Banach space is a bounded linear operator from the space in question to  $\mathbb{R}$ . Recall that in  $\text{RCA}_0$ , every bounded linear functional on a Banach space is equivalent to a linear continuous function. If  $f$  is a bounded linear functional,  $\ker f$  is a closed linear set, since it is the inverse image of the closed set  $\{0\}$ . It was mentioned above that it is also a closed subspace, provably in  $\text{RCA}_0$ , and that it is located if and only if the norm of  $f$  exists. These facts are proved in the following section.

Unless specified otherwise, assume that  $B$  is the underlying Banach space, spanned by the countable set  $A = \langle x_n \rangle$ .

**Proposition 2.3.1** ( $\text{RCA}_0$ ) *The kernel of a bounded linear functional in a Banach space is a closed subspace of that space.*

*Proof.* If  $f$  is the constant zero function, the statement is trivial. Thus it can be assumed that there is a  $y \in B$  such that  $f(y) \neq 0$ . Replacing  $y$  by  $y/f(y)$  if necessary, consider  $y$  such that  $f(y) = 1$ .

Define the sequence  $\langle a_n \rangle$  by  $a_n = x_n - f(x_n)y$ . Then for each  $n$

$$f(a_n) = f(x_n - f(x_n)y) = f(x_n) - f(x_n)f(y) = 0,$$

so  $a_n$  is in the kernel of  $f$ . It suffices to show that the sequence  $a_n$  is dense in the kernel, that is, for every  $n$  and every  $x$  such that  $f(x) = 0$ , exists  $x_m \in A$  such that  $\|x - (x_m - f(x_m)y)\| < 2^{-n}$ .

Because  $A$  is dense in  $B$ , choose  $x_m \in A$  such that  $\|x - x_m\| < \varepsilon$  (which will be specified later). Then, since  $f$  is bounded,  $|f(x_m)| \leq M\varepsilon$  (for all  $x \in B$ ,  $|f(x)| \leq M\|x\|$ ). Therefore,

$$\|x - (x_m - f(x_m)y)\| \leq \|x - x_m\| + |f(x_m)|\|y\| \leq \varepsilon + M\|y\|\varepsilon.$$

The choice for  $\varepsilon$  is now clear:  $\varepsilon = \min\{\frac{1}{2^{n+1}}, \frac{1}{2^{n+1}M\|y\|}\}$ . □

When is the kernel of a functional located? The answer to this question is provided by the next theorem. Its proof has been adapted from [8].



**Theorem 2.3.2** (RCA<sub>0</sub>) *Let  $f$  be a bounded linear functional. The following are equivalent:*

1.  $f$  has a norm.
2.  $\ker f$  is located.

*Proof.* (1)  $\rightarrow$  (2): Let  $f$  be a bounded linear functional whose norm is known. If  $f \equiv 0$  then  $d(x, \ker f) = 0$  for all  $x$ , so suppose  $f$  is not identically 0, in which case  $\|f\| > 0$ . We will show that for every  $x \in B$ ,

$$d(x, \ker f) = \frac{|f(x)|}{\|f\|}.$$

On the one hand, if  $f(z) = 0$ ,

$$|f(x)| = |f(x - z)| \leq \|f\|(\|x - z\|),$$

and  $\|x - z\| \geq \frac{|f(x)|}{\|f\|}$  so  $d(x, \ker f) \geq \frac{|f(x)|}{\|f\|}$ .

On the other hand, since  $\|f\| > 0$ , there is  $\varepsilon'$  such that  $\|f\| > \varepsilon'$ . Because  $f$  is normable, the quantity

$$\|f\| = \sup\left\{\frac{|f(x)|}{\|x\|} \mid x \in H\right\} = \sup\left\{\frac{|f(x)|}{\|x\|} \mid x \in A\right\}$$

exists. Fix  $\varepsilon < \varepsilon'$ . By definition of supremum, there exists  $y \in A$  such that

$$|f(y)| > (\|f\| - \varepsilon)\|y\|.$$

Let  $z = x - \frac{f(x)y}{f(y)}$ . Then,  $f(z) = 0$  and

$$\|x - z\| = \frac{|f(x)|\|y\|}{|f(y)|} < \frac{|f(x)|\|y\|}{(\|f\| - \varepsilon)\|y\|} = \frac{|f(x)|}{\|f\| - \varepsilon}.$$

Since  $\varepsilon$  can be made arbitrarily small,  $\frac{|f(x)|}{\|f\|}$  is an upper bound for the distance:  $d(x, \ker f) = \inf\{d(x, z) \mid f(z) = 0\} = \frac{|f(x)|}{\|f\|}$  and it is a continuous function, so the kernel is located.

As for (2)  $\rightarrow$  (1), because  $f$  is nonzero, there exists  $x_0$  such that  $f(x_0) = 1$ , and, as  $f$  is continuous,  $d(x_0, \ker f) > 0$ .

$$d(x_0, \ker f) = \inf\{\|x_0 - z\| \mid z \in \ker f\} = \inf\{\|y\| \mid f(y) = 1\}.$$

(If  $y = x_0 - z$ , then  $f(y) = f(x_0) - f(z) = 1$ ; on the other hand, if  $f(y) = 1$ , it can be written as  $y = x_0 - z$ , as in the proof of (1)  $\rightarrow$  (2) above.)

However,

$$\begin{aligned}\|f\| &= \sup\{|f(x)| \mid \|x\| = 1\} \\ &= \sup\{\|y\|^{-1} \mid y = \frac{x}{f(x)}, \|x\| = 1, f(x) \neq 0\} = \\ &= \sup\{\|y\|^{-1} \mid f(y) = 1\} = d(x_0, \ker f)^{-1},\end{aligned}$$

and, as the kernel is located, the norm of  $f$  exists.  $\square$

For Hilbert spaces the following theorem holds. Note that clause 3 is one form of the Riesz Representation Theorem.

**Theorem 2.3.3 (RCA<sub>0</sub>)** *Let  $f$  be a bounded linear functional on a Hilbert space  $H$ . The following are equivalent:*

1.  $f$  has a norm.
2.  $\ker f$  is located.
3. There is a  $y$  in  $H$  such that for every  $x$ ,  $f(x) = \langle x, y \rangle$ .

*Proof.* The equivalence of statements 1 and 2 has just been proved. It remains to show (2)  $\rightarrow$  (3) and (3)  $\rightarrow$  (1).

(2)  $\rightarrow$  (3): If  $f \equiv 0$ , simply take  $y = 0$ . Otherwise, there is an  $x_0$  such that  $f(x_0) \neq 0$ . By Proposition 2.2.9  $\ker f$  is a closed subspace of  $H$ , and by Theorem 2.2.7 the projection  $Px_0$  of  $x_0$  on  $\ker f$  exists. Let  $z = x_0 - Px_0$ , and recall from Lemma 2.2.6 that  $\langle x, z \rangle = 0$  whenever  $f(x) = 0$ . Finally, let  $y = \frac{f(z)z}{\|z\|^2}$ . We will show that this  $y$  has the required property.

Note that every  $x \in H$  can be written as  $x = u + \alpha y$ , where  $f(u) = 0$  and  $\alpha \in \mathbb{R}$ , since

$$x = \left(x - \frac{f(x)}{f(y)}y\right) + \frac{f(x)}{f(y)}y$$

and

$$f\left(x - \frac{f(x)}{f(y)}y\right) = f(x) - f(x) = 0.$$

Then,

$$\begin{aligned}\langle x, y \rangle &= \langle u + \alpha y, y \rangle = \langle u, y \rangle + \alpha \|y\|^2 \\ &= \alpha \|y\|^2 = \alpha \frac{|f(z)|^2 \|z\|^2}{\|z\|^4} \\ &= \alpha \frac{|f(z)|^2}{\|z\|^2}\end{aligned}$$

and

$$f(x) = x(u + \alpha y) = f(u) + \alpha f(y) = \alpha f(y) = \alpha \frac{(f(z))^2}{\|z\|^2},$$

so for all  $x \in H$ ,  $f(x) = \langle x, y \rangle$ . It is not hard to show that this  $y$  is unique.

(3)  $\rightarrow$  (1). If  $f(x) = \langle x, y \rangle$  for every  $x$ , it is straightforward to show that  $\|f\| = \|y\|$ .  $\square$

The following theorem is the uniform version of the previous one and also holds particularly for Hilbert spaces.

**Theorem 2.3.4** (RCA<sub>0</sub>) *The following are equivalent:*

1. Every bounded linear functional has a norm.
2. For every bounded linear functional  $f$ ,  $\ker f$  is located.
3. Every bounded linear functional is representable ( $\exists y \forall x f(x) = \langle x, y \rangle$ ).
4. (ACA)

*Proof.* By the preceding theorem, 1–3 are equivalent. Thus it suffices to show that 4 implies 1, and that 3 implies 4.

(4)  $\rightarrow$  (1):  $\|f\| = \sup\{\frac{|f(x)|}{\|x\|} \mid x \in A\}$ . As  $\{\frac{|f(x)|}{\|x\|} \mid x \in A\}$  is a bounded sequence of real numbers it has a least upper bound in ACA<sub>0</sub>.

(3)  $\rightarrow$  (4). Let  $\langle a_n \rangle$  be a sequence of real numbers; we may assume that for all  $n$ ,  $0 \leq a_n \leq 1$ . Next, let  $H$  be  $l^2$  (as defined in [26]).  $H$  has an orthonormal basis:  $e_n = \langle \delta_{n,m} \rangle$ .

Let  $b_i = \sqrt{a_{i+1} - a_i}$ , and define a functional  $f$  on the orthonormal basis by  $f(e_i) = b_i$  and extend linearly to all of  $H$ . Show that it is bounded.

For each fixed  $n$ ,

$$|f(\sum_{i=1}^n x_i e_i)| = |\sum_{i=1}^n x_i b_i| \leq (\sum_{i=1}^n x_i^2)^{\frac{1}{2}} (\sum_{i=1}^n b_i^2)^{\frac{1}{2}}.$$

The last inequality uses Hölder's inequality for sums in  $\mathbb{R}^n$ , which is easy to formalize in RCA<sub>0</sub>. Since

$$\sum_{i=1}^n b_i^2 = a_n \leq 1$$

for all  $n$ ,

$$|f(\sum_{i=1}^n x_i e_i)| \leq (\sum_{i=1}^n x_i^2)^{\frac{1}{2}} \leq \|x\| < \infty.$$

In particular, for every  $x \in A$  (the space of all finite sequences of rational numbers;  $l^2$  is the completion of  $A$  under the  $l^2$  norm),  $f(x)$  is well-defined,  $f$  is linear, and  $|f(x)| \leq \|x\|$  so  $f$  is a bounded linear functional.

By assumption, the Riesz representation theorem holds. Let  $y$  be the vector guaranteed to exist by the theorem. Then  $b_i = f(e_i) = \langle y, e_i \rangle = y_i$ . Because the theorem also tells us that  $\|y\|$  exists, since  $\|y\|^2 = \lim_n a_n$ , the theorem is proved.  $\square$

In case of a Banach space the following holds.

**Corollary 2.3.5** ( $\text{RCA}_0$ ) *The following are equivalent:*

1. *Every bounded linear functional on a Banach space is normable.*
2. *For every bounded linear functional  $f$ ,  $\ker f$  is located.*
3. *(ACA).*

## Chapter 3

# The Mean Ergodic Theorem

### 3.1 Proof of the Main Theorem

The mean ergodic theorem was initially stated and proved (by Von Neumann) for the space  $L_2(X)$  of square integrable functions. In this form, the theorem was closely related to Birkhoff's pointwise ergodic theorem.

In the original statement of the mean ergodic theorem, the transformation whose average effect on the system is examined acts on the space, i.e. the measure preserving transformation  $U$  is defined on the measure space  $X$ . This transformation induces a transformation  $T$  on  $L_2(X)$ , defined as  $(Tf)(x) = f(U(x))$  (see [16]). The induced transformation is an isometry on  $L_2(X)$ . This motivated Riesz, who was the first to consider a more general case, to replace  $L_2(X)$  with an arbitrary Hilbert space and to consider an isometry on that space. It is in this form that the mean ergodic theorem is stated and proved in standard sources today.

The statement of the theorem is:

Let  $T$  be an isometry of a Hilbert space, let  $x$  be any point, and consider the sequence  $\langle S_n x \rangle$  given by

$$S_n x = \frac{1}{n} (x + Tx + \dots + T^{n-1}x).$$

Then the sequence converges.

The classical proof of the mean ergodic theorem hinges on the fact that the Hilbert space can be represented as a direct sum of two subspaces. Given  $T$  as above, let  $M = \{x \mid Tx = x\}$  be the set of fixed points, and let  $N$  be the closure of the set  $\{Tx - x \mid x \in H\}$ . Classically,  $M$  and  $N$  are closed

subspaces and each other's orthogonal complements, so  $H = M \oplus N$ . The proof of the mean ergodic theorem examines the behavior of elements in both subspaces with respect to  $T$  and shows that  $S_n x$ , as described above, converges to the projection of  $x$  onto  $M$ .

Some of the difficulties connected to proving the mean ergodic theorem in weak subsystems of second-order arithmetic were mentioned in Section 1.3. We will now continue that discussion. A number of interesting problems arise when one tries to translate all these facts into our framework. There is no suitable definition for closure of a set so  $N$  is represented with the code  $\langle x - Tx \mid x \in A \rangle$ , with the understanding that this is a code for a closed subspace of  $H$  and every point of  $N$  is represented with a Cauchy sequence  $\langle x_n - Tx_n \rangle$ . Since  $A$  is dense in  $H$ , the set thus defined is precisely  $N$ . Next, as was mentioned before, one can only conclude that  $M$  is a closed linear set (it is the kernel of the continuous function  $f(x) = \|Tx - x\|$ ), not that it is a closed subspace. Similarly, it cannot be deduced that  $N$  is a closed set. Finally, though it is not difficult to show that  $M = N^\perp$  (this will be done below), it is not possible in  $\text{RCA}_0$  to show the converse to this statement. In order to show  $x \perp M \rightarrow x \in N$ , one needs to exhibit an element of  $N$  that  $x$  is equal to, which may not be possible.

The main result is that the mean ergodic theorem is equivalent to  $(ACA)$ :

**Theorem 3.1.1** ( $\text{RCA}_0$ ) *The following statements are equivalent:*

1. *For every Hilbert space  $H$ , nonexpansive linear operator  $T$ , and point  $x$ , the sequence of partial averages  $\langle S_n x \rangle$  converges.*
2. *For every Hilbert space  $H$ , isometry  $T$ , and point  $x$ , the sequence of partial averages  $\langle S_n x \rangle$  converges.*
3.  $(ACA)$ .

Clearly 1 implies 2. That 2 implies 3 is proved in the next section, and 3 implies 1 is Corollary 3.1.3.

The proof of the statement presented below is constructive. Significant parts of it were adapted from [27]. It establishes the mean ergodic theorem and gives a finer analysis of conditions needed for it to hold.

**Theorem 3.1.2** ( $\text{RCA}_0$ ) *Let  $T$  be any nonexpansive linear operator on a Hilbert space and let  $x$  be any point. With the notation above, the following are provably equivalent in  $\text{RCA}_0$ :*

1.  $P_N x$  exists.

2.  $x$  can be written as  $x = x_M + x_N$ , where  $x_M \in M$  and  $x_N \in N$ .

3.  $\lim_n S_n(x)$  exists.

Furthermore, if these statements hold, then the decomposition in 2 is unique and  $P_M x$  also exists. In fact, the following equalities hold:

$$\lim_n S_n(x) = P_M x = x_M = x - P_N x.$$

*Proof.* First, let us show that  $M \cap N = \{0\}$ . If  $x \in M \cap N$ , then  $Tx = x$  (and so  $S_n x = x$  for all  $n$ ), and for every  $\varepsilon$  there is some  $y$  such that  $\|x - (y - Ty)\| < \varepsilon$ , which implies that  $\|S_n x - S_n(y - Ty)\| = \|x - S_n(y - Ty)\| < \varepsilon$ . But  $\|S_n(y - Ty)\| \rightarrow 0$  (proved in more detail below) so  $\|x\| < \varepsilon$  and consequently,  $\|x\| = 0$ .

Next, show that if an element  $y \in H$  is orthogonal to  $N$ , it is in  $M$ . Let  $y \perp N$ . Then  $\langle y, x - Tx \rangle = 0$  for all  $x \in H$ . In particular,  $\langle y, y - Ty \rangle = 0$ , or

$$\langle y, Ty \rangle = \langle y, y \rangle = \|y\|^2,$$

so

$$\begin{aligned} \|Ty - y\|^2 &= \langle Ty - y, Ty - y \rangle = \|Ty\|^2 - 2\langle Ty, y \rangle + \|y\|^2 \\ &\leq \|y\|^2 - 2\|y\|^2 + \|y\|^2 = 0. \end{aligned}$$

Therefore,  $y = Ty$ , and  $y \in M$ , as required.

Let us now prove the three statements above equivalent.

(1)  $\rightarrow$  (2): Write  $x = (x - P_N x) + P_N x$ . Clearly,  $P_N x$  is in  $N$ . It remains to show that  $x - P_N x \in M$ . Since  $x - P_N x$  is orthogonal to  $N$ , the argument above shows that  $x - P_N x \in M$ , as required. Note that the fact that  $M \cap N = \{0\}$  implies moreover that this decomposition is unique.

(2)  $\rightarrow$  (3): If  $x \in H$  and  $x = x_M + x_N$  then  $S_n(x_M) = x_M$  for all  $n$ , so  $\lim_n S_n(x_M) = x_M$ . Since  $x_N \in N$ , for every  $m$  exists  $u_m \in A$  such that

$$\|x_N - (u_m - Tu_m)\| \leq 2^{-m}.$$

Then

$$S_n(u_m - Tu_m) = \frac{1}{n} \sum_{k=1}^{n-1} (T^{k-1}u_m - T^k u_m) = \frac{1}{n} (u_m - T^n u_m)$$

and

$$\|S_n(u_m - Tu_m)\| \leq \frac{1}{n} (\|u_m\| + \|T^n u_m\|) \leq \frac{2\|u_m\|}{n} \rightarrow 0.$$

Also,

$$\begin{aligned} \|S_n x_N - S_n(u_m - Tu_m)\| &= \|S_n(x_N - (u_m - Tu_m))\| \leq \\ &\leq \|x_N - (u_m - Tu_m)\| \\ &\leq 2^{-m}, \end{aligned}$$

and therefore  $S_n(x_N) \rightarrow 0$  as well. Finally,  $S_n(x) = S_n(x_M) + S_n(x_N) \rightarrow x_M + 0 = x_M$ .

(3)  $\rightarrow$  (1): If  $\lim_n S_n(x)$  exists, then we expect it to be  $x - P_N(x)$ , so define  $P_N x : A \rightarrow H$  by

$$P_N(x) = x - \lim_n S_n(x)$$

and show that this works.

Use Lemma 2.2.6. Let  $\lim_n S_n(x) = y$ . We need to show that  $x - y \in N$  and that  $x - (x - y) = y$  is orthogonal to  $N$ .

Let  $y_n = (\frac{n-1}{n}I + \frac{n-2}{n}T + \dots + \frac{1}{n}T^{n-1})x$ . Note that  $(I - T)y_n = (I - S_n)x \rightarrow x - y$  which shows that  $x - y$  is in  $N$ . For the second part, first show that  $y \in M$ , i.e.  $Ty = y$ , which is true by the following argument:

$$TS_n x = \frac{1}{n}(Tx + \dots + T^n x) = S_n x + \frac{1}{n}(T^n x - x),$$

and  $\lim_n TS_n x = \lim S_n x$  or  $Ty = y$ . Since  $y \in M$  if  $z \in N$ ,  $\langle y, z \rangle = 0$ .  $\square$

**Corollary 3.1.3** (ACA<sub>0</sub>) *The mean ergodic theorem holds, i.e. for every Hilbert space  $H$  and nonexpansive mapping  $T$ , for every  $x \in H$ ,  $\langle S_n x \rangle$  converges.*

*Proof.* Since  $N$  is a closed subspace, by Theorem 2.2.7 ACA<sub>0</sub> proves that  $P_N$  exists. It was shown above that for all  $x$ ,  $S_n x$  converges to  $(I - P_N)x$ .  $\square$

Just as knowing that  $N = M^\perp$  does not help us conclude that  $M = N^\perp$ , the hypothesis that  $P_M$  exists provides no help at all in proving the mean ergodic theorem. These two facts are closely related. If  $P_N$  exists, it is provable in RCA<sub>0</sub> that  $P_M = I - P_N$ . The other direction, however, is not true. Therefore, knowledge of  $P_M$  does not provide us with  $P_N$ , or with the decomposition in the second item of the previous theorem, or for that matter, with the proof of convergence of  $\langle S_n x \rangle$ , even though the limit, if it exists, is equal to  $P_M x$ .



## 3.2 Reversal

This section will provide the conclusion to the proof of Theorem 3.1.1 and tie some other loose ends.

We will begin by listing two direct and simple reversals: one for the case of nonexpansive transformation  $T$  and the other for an isometry. Naturally, the second reversal would suffice, but both are given for illustrative purposes.

First suppose  $\langle a_i \rangle$  is a sequence of reals in  $[0, 1]$ , and let  $H$  be the space  $l_2$ . Define an operator  $T$  on  $l_2$  by  $T(e_i) = (1 - a_i)e_i$ . Then  $T$  is a nonexpansive mapping because  $0 \leq a_n \leq 1$  for every  $n$ , and  $S_n e_i = e_i$  if  $a_i = 0$ , while

$$S_n e_i = \frac{1}{n} \sum_{m=0}^{n-1} (1 - a_i)^m e_i = \frac{1}{n} \cdot \frac{1 - (1 - a_i)^n}{a_i} e_i$$

otherwise, which clearly converges to 0. Let  $x = \sum_i (1/2^i) e_i$ , and let  $y = \lim_n S_n x$ . Then for each  $i$ ,  $\langle y, e_i \rangle \neq 0$  if and only if  $a_i = 0$ , providing a  $\Sigma_1$  equivalent to the  $\Pi_1$  assertion  $a_i = 0$ . By Lemma 1.2.3, this shows that the mean ergodic theorem for nonexpansive mappings implies arithmetic comprehension.

The reversal when  $T$  is an isometry is somewhat more involved. If we still want to consider square-summable sequences and use a similar argument, the question arises of how to define  $T$  so that it preserves norm. If  $l_2$  is treated as a real Hilbert space, this cannot be done. Instead, consider  $l_2(\mathbb{C})$ .

Given any sequence of real numbers  $a_n$  in  $[0, 1]$ , define the sequence of complex numbers

$$z_n = \frac{1 + a_n i}{|1 + a_n i|}.$$

So  $|z_n| = 1$  for each  $n$  and  $z_n = 1$  if and only if  $a_n = 0$ . Define a linear operator  $T$  on  $l_2(\mathbb{C})$  by  $T(e_k) = z_k e_k$ . The fact that  $|z_k| = 1$  for each  $k$  implies that  $T$  is an isometry. If  $z_k = 1$ , then  $S_n e_k = e_k$ ; otherwise,

$$\begin{aligned} S_n e_k &= \frac{1}{n} (1 + z_k + z_k^2 + \dots + z_k^{n-1}) e_k \\ &= \frac{1 - z_k^n}{n(1 - z_k)} e_k, \end{aligned}$$

which converges to 0 as  $n$  increases, since

$$\begin{aligned} \left| \frac{1 - z_k^n}{n(1 - z_k)} \right| &\leq \frac{|1| + |z_k^n|}{n|1 - z_k|} \\ &= 2/(n|1 - z_k|). \end{aligned}$$

As above, if  $x = \sum_k (1/2^k)e_k$  and  $y = \lim_n S_n x$ , then  $\langle y, e_k \rangle \neq 0$  if and only if  $z_k = 0$ , i.e. if and only if  $a_k = 0$ . Once again, by Lemma 1.2.3, this implies arithmetic comprehension.

The results stated below provide strengthenings for the reversal. The remainder of the section contains results proved for the most part by Jeremy Avigad, presented here for completeness. The proof of the next theorem and the corollary that follows it has been taken from [2], while 3.2.3 is stated without proof.

**Theorem 3.2.1** ( $\text{RCA}_0$ ) *Each of the following statements is equivalent to (ACA):*

1. *For any nonexpansive mapping  $T$  on a Hilbert space,  $M$  is a closed subspace.*
2. *For any isometry  $T$  on a Hilbert space,  $M$  is a closed subspace.*
3. *For any nonexpansive mapping  $T$  on a Hilbert space,  $N$  is a closed linear set.*
4. *For any isometry  $T$  on a Hilbert space,  $N$  is a closed linear set.*
5. *For any nonexpansive mapping  $T$  on a Hilbert space and any  $x$ , the projection  $P_M x$  exists.*
6. *For any isometry  $T$  on a Hilbert space and any  $x$ , the projection  $P_M x$  exists.*
7. *For any nonexpansive mapping  $T$  on a Hilbert space and any  $x$ , the distance from  $x$  to  $M$  exists.*
8. *For any isometry  $T$  on a Hilbert space and any  $x$ , the distance from  $x$  to  $M$  exists.*

*Proof.* By Corollary 3.1.3, (ACA) implies that both  $P_M$  and  $P_N$  exist for any nonexpansive mapping  $T$ . This implies 5 right away. Also, by Theorems 2.2.8 and 2.2.7 respectively, it implies that  $M$  and  $N$  are both located; and since under the assumption of locatedness, a set is closed if and only if it is separably closed, this, in turn, implies that  $M$  and  $N$  are separably closed and closed, respectively. Hence (ACA) proves 1, 3, and 5. Clearly 1 implies 2, 3 implies 4, 5 implies 6–8, 6 implies 8, and 7 implies 8. Thus we only need to establish reversals from each of 2, 4, and 8 to (ACA).

Let us first show that 2 implies (ACA). We will use statement 3. in Lemma 1.2.3. Given a sequence of real numbers  $\langle a_k \mid k \in \mathbb{N} \rangle$  in  $[0, 1]$ , define an isometry  $T$  as described just before the statement of the lemma. By 2, the set  $M = \{x \mid Tx = x\}$  is a closed subspace, with a countable dense sequence  $\langle y_k \mid k \in \mathbb{N} \rangle$ . But then for every  $k$  we have

$$\begin{aligned} a_k = 0 &\leftrightarrow e_k \in M \\ &\leftrightarrow \exists j (\|e_k - y_j\| < 1), \end{aligned}$$

providing a  $\Sigma_1$  definition of  $\{k \mid a_k = 0\}$ . (To see that  $\exists j (\|e_k - y_j\| < 1)$  implies that  $e_k \in M$ , note that if  $e_k \notin M$  we have  $\langle e_k, y_j \rangle = 0$  for every  $y_j$ . But then  $\|e_k - y_j\|^2 = \langle e_k - y_j, e_k - y_j \rangle = \|e_k\|^2 + \|y_j\|^2 = 1 + \|y_j\|^2 \geq 1$ .)

To show 4 implies (ACA), given  $\langle a_k \rangle$  we use the same  $T$ . Assuming 4,  $N$  is a closed set. But then  $a_k = 0 \leftrightarrow e_k \notin N$  again provides a  $\Sigma_1$  definition of  $\{k \mid a_k = 0\}$ .

The reversal from 8 to (ACA) is omitted and can be found in Section 15 of [2].  $\square$

**Corollary 3.2.2** ( $\text{RCA}_0$ ) *Each of the following implies (ACA):*

1. *If  $S$  is any closed subspace of a Hilbert space, then  $S$  is a closed linear set.*
2. *If  $S$  is any closed linear set in a Hilbert space, then  $S$  is closed subspace.*
3. *If  $S$  is any closed linear set in a Hilbert space and  $x$  is any point, the projection of  $x$  on  $S$  exists.*
4. *If  $S$  is any closed linear set in a Hilbert space and  $x$  is any point, the distance from  $x$  to  $S$  exists.*

The second implication completes the proof of Theorem 2.2.2. The third and fourth implications complete the proof of Theorem 2.2.11.

**Theorem 3.2.3**  $\text{RCA}_0$  *proves that the following are equivalent:*

1. *For every  $x$  in a Hilbert space and isometry  $T$ , if  $P_M x$  exists, then  $\lim_n S_n x$  exists.*
2. *For every  $x$  in a Hilbert space and isometry  $T$ , if  $P_M x = 0$ , then  $\lim_n S_n x$  exists.*

3. For every  $x$  in a Hilbert space and isometry  $T$ , if  $P_M x = 0$ , then  $\lim_n S_n x = 0$ .
4. (ACA)

Of the 4 statements listed, statement 4 implies 1 by Corollary 3.1.3, clearly 1 implies 2, and 2 implies 3 by Theorem 3.1.2. So, the only part remaining to be shown is that 3 implies 4. The construction is based on a strategy employed in a different context by Pour-el and Richards ([22]). Once again, for details see [2].

## Chapter 4

# Aspects of Measure Theory

### 4.1 Preliminaries

The classical theory of measure is highly nonconstructive. Bishop and Bridges state in [8]:

Any constructive approach to mathematics will find a crucial test in its ability to assimilate the intricate body of mathematical thought called measure theory, or the theory of integration.

Simpson writes in ([26], X.1):

Historically, the subject of measure theory developed hand in hand with the nonconstructive, set-theoretic approach to mathematics... Although Reverse Mathematics is quite different from Bishop-style constructivism, we see that Bishop's remark raises an interesting question: *Which nonconstructive set existence axioms are needed for measure theory?*

In their work in measure theory, Simpson and his students (principally Yu), have utilized an approach similar to Bishop's, modifying the axioms for the Daniell integral. The basic objects are functions, not sets. Just as metric, Banach and Hilbert spaces are presented as completions of certain countable sets, the space of integrable functions is given as the completion of the space of simple functions under the  $L_1$  norm: it is, in fact, a Banach space. All spaces considered are compact probability spaces. Every measurable set is identified with its characteristic function, so a set is measurable if and only if the characteristic function is measurable. In order to remain consistent with the standard practice of measure theory, I am going to diverge from

the terminology in [26] and call functions in  $L_1(X)$  *integrable*, not measurable. Similarly, a set whose characteristic function is integrable will itself be called an integrable set. The classical notion of a measurable function will not occur in what follows. Classically, a measurable function that is not integrable has infinite measure, which will never be the case in our setup.

$\text{ACA}_0$  is a natural system to work with in measure theory, as it provides us with least upper bounds and greatest lower bounds of sequences, assuring existence of measure for countable unions and intersections of integrable sets. However,  $\text{WWKL}_0$  turns out to be sufficient for a large portion of discourse about integrable functions, as it proves some of their relevant pointwise properties. Yu points this out in [31]:

These results are expected, if we recall that Littlewood’s three principles are centered at the idea of “nearly.” . . . the weak-weak König’s lemma provides means of doing things “nearly” continuous. The weak version of Heine-Borel theorem, which is equivalent to weak-weak König’s lemma, gives finite “subcovers” with small errors for any open covering. Hence it is reasonable to expect to develop the theory of integration in the subsystem  $\text{WWKL}_0$ .

If one is careful with definitions, it is possible to talk about these things even in  $\text{RCA}_0$ . It is usually possible to formalize standard definitions or find equivalent ones in  $\text{RCA}_0$ . Though it can sometimes be shown that desired properties of these definitions hold in  $\text{RCA}_0$ , those proofs are usually more complicated and circuitous. Oftentimes it is necessary to reason about pointwise properties of functions, and in those cases stronger axioms, like ( $\text{WWKL}$ ) or ( $\text{ACA}$ ) are employed.

Relevant definitions and facts about measures are listed below. More information can be found in [32, 31, 26, 11]. These sources provide an in-depth treatment of measure theory in second-order arithmetic. Unless specified otherwise, all definitions are stated in  $\text{RCA}_0$ .

The Banach space  $C(X)$ , elements of which are functions with a modulus of uniform continuity, is the completion of the countable set  $S(X)$  of simple functions over  $X$  under the sup norm. More precisely:

**Definition 4.1.1** *Let  $X = \hat{A}$  and  $d$  a metric on  $X$ . Let  $B = A \times \mathbb{Q}^+ \times \mathbb{Q}^+$ . Each  $b = (a, r, s) \in B$  is intended to represent a function  $\phi_b$  as follows:*

$$\phi_b(x) = \begin{cases} 1 & d(a, x) \leq x \\ \frac{r-d(a,x)}{r-s} & s < d(a, x) < r \\ 0 & d(a, x) \geq r. \end{cases}$$

Define  $C = \mathbb{Q} \times B$ . If  $c = \{(q_1, b_1), \dots, (q_n, b_n)\}$  is a finite subset of  $C$ , define  $\phi_c : X \rightarrow \mathbb{R}$  by  $\phi_c(x) = \sum_{k=1}^n q_k \phi_{b_k}(x)$ .

Let  $S(X) = \{c \mid c \text{ is a finite subset of } C\}$ . Finally,  $C(X) = \hat{S}(X)$  under the sup norm given by  $\|c\| = \sup_{x \in X} |\phi_c(x)|$ .

The elements of  $S(X)$  will be referred to as *simple functions*, disregarding the distinction between functions and their codes, as is customary when working in second-order arithmetic. Bishop and Bridges refer to elements of  $C(X)$  as *test functions*, which is terminology we will also occasionally use. Notice that test functions are bounded, i.e. if  $f \in C(X)$ , then there is some constant  $M_f \in \mathbb{R}$  such that  $|f(x)| \leq M_f$  for all  $x$ .

**Definition 4.1.2** *If  $X$  is a compact metric space, a Borel measure  $\mu$  is a nonnegative bounded linear functional  $\mu : C(X) \rightarrow \mathbb{R}$  such that  $\mu(1) = 1$ .*

In the case  $X = [0, 1]$ , a natural interpretation for  $\mu$  is the Lebesgue measure, where for  $f \in C([0, 1])$ ,  $\mu(f) = \int_0^1 f(x) dx$ . Measure of open and closed sets is defined via test functions.

**Definition 4.1.3** *If  $U$  is an open set, and  $\mu$  a Borel measure on  $C(X)$ , then define*

$$\mu(U) = \sup\{\mu(g) \mid g \in C(X), 0 \leq g \leq 1, g = 0 \text{ on } X \setminus U\}.$$

*The definition in the case of closed sets is dual:  $\mu(C) = \inf\{\mu(h) \mid h \in C(X), 0 \leq h \leq 1, h = 1 \text{ on } C\}$ .*

These suprema may not exist; in fact, it is not difficult to show that the statement that every open set has a measure is equivalent to (ACA). More properties of this measure can be found in [31]. For example, it is provable in  $\text{RCA}_0$  that it is regular on open and closed sets.

The definition of almost everywhere (a.e.) convergence, which will be used in the statement of the pointwise ergodic theorem, requires the concept of a *null  $G_\delta$  set*. A  $G_\delta$  set is coded by a sequence of open sets (meaning that it is the intersection of this sequence). It can be assumed that this sequence is decreasing. Let  $G$  be a  $G_\delta$  set coded with a sequence  $\langle U_n \rangle$ . If  $\mu(U_n) \leq 2^{-n}$  for all  $n$ , then  $G$  is a *null  $G_\delta$  set*. The complement of a null  $G_\delta$  set is a *full  $F_\sigma$  set*. Almost everywhere convergence is defined as convergence outside of a null  $G_\delta$  set (or equivalently, convergence on a full  $F_\sigma$  set).

**Definition 4.1.4** *Let  $p \geq 1$  be a real number and  $X$  a compact metric space. The separable Banach space  $L_p(X)$  is the completion of  $S(X)$  (the*

space of simple functions) under the  $L_p$  norm:  $\|f\|_p = \mu(|f|^p)^{1/p}$ . A function  $f \in L_p(X)$  is a strong Cauchy sequence with respect to the norm of functions in  $S(X)$ . In other words,  $f = \langle f_n \mid n \in \mathbb{N} \rangle$  and  $\|f_m - f_n\|_p \leq 2^{-m}$  for  $n > m$ .

Notice that the definition implies that if  $\langle f_n \rangle$  is a representation of  $f \in L_p(X)$ , then  $\|f_n\|_p \rightarrow \|f\|_p$  in  $\mathbb{R}$ .

The definition works for all real  $p > 1$ , though this fact is not obvious: it is not obvious how to define  $|f|^p$  when  $p$  is not a natural number, but it can be shown, using the definition of power functions given in Appendix B, that if  $f \in S(X)$ ,  $|f|^p \in C(X)$  so the definition above makes sense. The case of special interest is when  $p = 1$ . The elements of  $L_1(X)$  are called integrable functions. We sometimes write  $\mu(f) = \int f d\mu$  and usually omit the subscript in the norm. It is important to point out the following result, which can be found in [30] and [11].

**Proposition 4.1.5 (WWKL<sub>0</sub>)** *If  $f$  is an element of  $L_1(X)$ , represented as  $\langle f_n \rangle$ , then  $f(x)$  is defined a.e. and is equal to  $\lim f_n(x)$ .*

It will later be shown that  $L_p(X) \subseteq L_1(X)$  for all  $p > 1$ , so elements of all  $L_p$  spaces are pointwise defined in WWKL<sub>0</sub>. It is important to keep in mind that functions in  $L_p(X)$  are elements of a Banach space. Their pointwise properties are not known in RCA<sub>0</sub> and  $f = g$  means, by definition, equality in the normed space:  $f = g \leftrightarrow \|f - g\|_p = 0$ .

*Note:* It is sometimes convenient to presuppose that if  $f \in L_p(X)$ , then  $\langle f_n \rangle$  is a sequence of test functions instead of simple functions.

If  $f, g \in L_p(X)$ , then it is possible to define  $\max(f, g)$  and  $\min(f, g)$ . The former is represented as  $la \max(f_n, g_n)$  and the latter as  $\langle \min(f_n, g_n) \rangle$  and it is easily shown that they are elements of the space  $L_p(X)$ . Furthermore, if  $f = f'$  and  $g = g'$ , then  $\max(f, g) = \max(f', g')$  and  $\min(f, g) = \min(f', g')$ . In particular, for  $f \in L_p(X)$ , we can define  $f^+ = \max(f, 0)$ , represented as  $\langle \max(f_n, 0) \rangle$ , and  $f^- = \max(-f, 0)$ , with the representation  $\langle \max(-f_n, 0) \rangle$ . It immediately follows (in RCA<sub>0</sub>) that  $f^+, f^- \in L_p(x)$  and  $f = f^+ - f^-$ . As  $|f| = f^+ + f^-$ ,  $|f|$  is integrable whenever  $f$  is.

The following definitions are substitutes in RCA<sub>0</sub> for the usual, pointwise definitions of properties in question.

**Definition 4.1.6 (RCA<sub>0</sub>)** *A function  $f \in L_p(X)$  is nonnegative if  $|f| = f$  in  $L_p(X)$ .*

*Note:* Equivalently,  $f$  is nonnegative if and only if  $f = f^+$  if and only if  $f^- = 0$ .



*Fact* [RCA<sub>0</sub>]: It can be assumed that if  $\langle f_n \rangle$  is a nonnegative function, then  $f_n(x) \geq 0$  for all  $n$  and all  $x$ , because otherwise  $f_n$  can be replaced by  $f_n^+$ .

*Fact* [WWKL<sub>0</sub>]: A function  $f$  is nonnegative in the sense defined above if and only if  $f(x) \geq 0$  a.e.

Similarly, the  $\geq$  relation between two functions in  $L_p(X)$  can be introduced in RCA<sub>0</sub>,

**Definition 4.1.7 (RCA<sub>0</sub>)** *Given two functions  $f$  and  $g$  in  $L_p(X)$ ,  $f \geq g$  if and only if  $\max(f, g) = f$ .*

*Note:* Equivalently,  $f \geq g$  if and only if  $\min(f, g) = g$ .

*Fact* [RCA<sub>0</sub>]:  $f \geq g$  if and only if  $|f - g| = f - g$  if and only if  $(f - g)^+ = (f - g)$  if and only if  $(f - g)^- = 0$  since  $f \geq g \leftrightarrow f - g \geq 0$ .

*Fact* [WWKL<sub>0</sub>]: The ordering defined above coincides with the usual, pointwise ordering a.e.

It is also possible to define  $f < g$  as  $\neg(f \geq g)$ .

To discuss products of functions it is necessary to consider a class of functions that is closed under products and powers: these are *essentially bounded* functions. They correspond to the classical space  $L_\infty$  whose elements are measurable functions such that for some  $M \in \mathbb{R}$ ,  $|f(x)| \leq M$  a.e. As there is no natural analogue of the concept of measurable function in reverse mathematics, here, as in a few other places, we will find a sensible substitute for the stipulation of measurability. In fact, for our purposes measurable functions can be avoided all together. In this particular case, instead of the entire space  $L_\infty(X)$ , we are only interested in the functions in  $L_p(X) \cap L_\infty(X)$  which can be characterized with the following definition:

**Definition 4.1.8 (RCA<sub>0</sub>)** *A function  $f \in L_p(X)$  is said to be essentially bounded if  $|f| \leq M$  for some  $M \in \mathbb{R}^+$ .*

*Note:* We think of  $M$  above as the constant function  $M$ , and then interpret  $|f| \leq M$  in the sense of the definition of the  $\leq$  relation.

Another convenient characterization is this:  $f$  is essentially bounded if and only if  $f = \max(\min(f, M), -M)$  for some  $M \in \mathbb{R}^+$ .

We are also not interested in considering  $L_{p, \text{infy}}(X)$  as normed spaces, since we only look at individual essentially bounded functions. It is worthwhile pointing out, however, that  $L_{1, \infty}$  corresponds to the classical  $L_\infty(X)$ , since  $X$  is a finite measure space, and for any  $p$ , every  $f \in L_{p, \text{infy}}(X)$  is in  $L_1(X)$ .

**Proposition 4.1.9** (RCA<sub>0</sub>) *The function  $f$  is essentially bounded if it is represented as  $\langle f_n \rangle$ , where each  $f_n$  is a simple function and there exists a constant  $M$  such that  $|f_n| \leq M$  for all  $n$ .*

*Proof.* Let  $f$  be represented with  $\langle f_n \rangle$ . Let  $M$  be the bound for  $f$ . Then  $f = \max(\min(f, M), -M)$ , and the right-hand side of this expression is represented as  $\langle \max(\min(f_n, M), -M) \rangle$ . This exactly means that every  $|f_n|$  is bounded by  $M$ , as required.  $\square$

*Fact:[RCA<sub>0</sub>]* If one representation of a function is essentially bounded, all representations are, and with the same bound, since if  $f = g$ , then  $\max(f, M) = \max(g, M)$  and  $\min(f, M) = \min(g, M)$ .

*Fact [WWKL<sub>0</sub>]:* The definition above implies the usual characterization of essential boundedness, since

$$\forall n |f_n| \leq M \rightarrow \forall n |f_n(x)| \leq M \rightarrow |f(x)| \leq M$$

for almost all  $x$ .

We will write “ $f \in L_{p,\infty}(X)$ ” to mean “ $f$  is an essentially bounded function in  $L_p(X)$ .”

Although this is a natural place to define products and analyze their integrability, the discussion about these issues is complex enough to merit a section of its own, and is therefore postponed until Section 4.2. The next order of business is instead to define characteristic functions and integrable sets, though this also needs some results from the next section. The definition found in [26] is made in WWKL<sub>0</sub> because it presupposes that functions are defined a.e. We first give this definition, and then show how it can be modified to make sense in RCA<sub>0</sub>.

**Definition 4.1.10** (WWKL<sub>0</sub>) *A function  $f \in L_1(X)$  is a characteristic function if  $f(x) \in \{0, 1\}$  a.e. A code for an integrable set  $E$  with respect to the measure  $\mu$  on  $X$  is defined to be a corresponding characteristic function and  $\mu(E) = \mu(f)$ .*

*Fact [WWKL<sub>0</sub>]:* A characteristic function defined in such a way is essentially bounded:  $0 \leq \chi_A \leq 1$ . Because of this,  $\chi_A^2 \in L_1(X)$  (this will be proved in Lemma 4.2.5); also,  $\chi_A^2(x) = \chi_A(x)$  for every  $x$  in the domain. Moreover, if  $f$  is any function such that  $f^2 = f$ , then  $f \geq 0$  and it follows that  $(1 - f)^2 = 1 - f$ , so  $1 - f \geq 0$ , or  $f \leq 1$ .

This motivates the following definition:

**Definition 4.1.11** ( $\text{RCA}_0$ ) *A function  $f \in L_1(X)$  is a characteristic function if  $f^2 = f$ .*

The definition of integrable set remains the same as in 4.1.11.

*Fact* [ $\text{WWKL}_0$ ]: The two definitions coincide.

*Proof.* One direction is the previous fact. The other follows from Proposition 4.4.1:  $f^2 = f$  if and only if  $f^2(x) = f(x)$  a.e. The latter means that there is an  $F_\sigma$  set  $F = \cup_n C_n$  on which  $f^2(x) = f(x)$ , which is equivalent to saying  $f(x) \in \{0, 1\}$  for all  $x \in F$ .

If  $E = (a, b)$  is an open interval in  $[0, 1]$ , its characteristic function can be coded with the sequence  $\langle p_n \rangle$  of simple functions, where  $p_n = \langle \frac{a+b}{2}, \frac{a-b}{2}, (1 - 2^{-n}) \frac{a-b}{2} \rangle$ .  $\text{RCA}_0$  proves that  $\langle p_n \rangle$  satisfies both definitions of a characteristic function, and  $\mu(E) = \lim_n \mu(p_n)$  (compare to [30]).

*Note:* We are not interested in general Borel sets, and for this reason entirely skip the definition and discussion of Borel codes. All the sets appearing in the theorems that follow are simple enough: they can be expressed with up to four quantifiers. Just as a  $G_\delta$  set can be coded as a sequence of open set, with the understanding that it is an intersection of these sets, so can, for example, an infinite union of  $G_\delta$  sets be coded as a sequence of sequences of open sets. A similar procedure applies for all sets we will be considering later.

Integrable sets are identified with their characteristic functions. When working in  $\text{WWKL}_0$ , where integrable functions are pointwise defined a.e, given a characteristic function  $f$ , we can define membership in the corresponding set  $A$ :  $x \in A \leftrightarrow f(x) = 1$ . Notice that there is no discrepancy between this definition and the previous definition of measure for open sets: in general, to define the characteristic function of an open set, arithmetic comprehension is necessary.

The complement  $\overline{A}$  of an integrable set  $A$  is integrable, since if  $f$  is a characteristic function of  $A$ , then a characteristic function of  $\overline{A}$  is  $1 - f$ . Similarly, if  $A$  and  $B$  are integrable sets with characteristic functions  $f_1$  and  $f_2$ , characteristic functions for  $A \cup B$  and  $A \cap B$  are respectively  $\max(f_1, f_2)$  and  $\min(f_1, f_2)$ . It is not difficult to show that the characteristic function for  $A \cap B$  can equivalently be written as  $f_1 \cdot f_2$  and the characteristic function for  $A \cup B$  as  $f_1 + f_2 - f_1 \cdot f_2$ . To show that these functions really correspond to complements, unions and intersections, ( $\text{WWKL}$ ) is needed. Finite unions and intersections can be shown to be integrable in  $\text{RCA}_0$ , but ( $\text{ACA}$ ) is needed for the infinite case. If  $f_n$  is the characteristic function of  $A_n$ , then

the characteristic function of  $\cup_n A_n$  is  $\sup_n f_n$  and the characteristic function of  $\cap_n A_n$  is  $\inf_n f_n$ . A proof that suprema and infima of integrable functions are integrable is provided in Proposition 4.5.4, and it is not difficult to show (in  $\text{WWKL}_0$ , provided they exist) that suprema and infima of characteristic functions are characteristic functions themselves. In particular, in  $\text{ACA}_0$ ,  $G_\delta$  and  $F_\sigma$  sets, as well as their unions and intersections, are integrable, or, for that matter any set at a very low level in the Borel hierarchy. If  $M$  is a Borel set of finite rank,  $\text{ACA}_0$  proves that there exists a  $G_\delta$  set  $G$  such that  $M \subseteq G$  and  $\mu(G \setminus M) = 0$ . This implies that every null set that is “simple enough”, that is, at a low level in the Borel hierarchy, is contained in a null  $G_\delta$  set. This will allow us to prove a.e. convergence of a sequence of functions even if the set on which the sequence converges is not  $F_\sigma$ . The fact follows immediately from Yu’s observation in [30] that  $\text{ACA}_0$  proves measure for Borel sets of finite rank to be well-defined and regular (this fact requires  $\text{ATR}_0$  in the general case). For, if  $\mu(M) = \inf\{\mu(U) \mid M \subseteq U \text{ and } U \text{ is open}\}$ , then there exists a sequence  $U_n$  of open sets such that  $M \subseteq U_n$  and  $\mu(U_n) \leq \mu(M) + 2^{-n}$  for all  $n$ . The set  $G = \cap_n U_n$  is the required  $G_\delta$  set.

In some cases it is possible to build a theory of measure and integration from sets. When  $X$  is  $[0, 1]^n$  or the Cantor space, there is a simple representation of measurable sets, as the completion of the Boolean algebra of clopen sets with respect to symmetric differences (see [11]). In this case, the definitions are made in  $\text{RCA}_0$  and the equivalence between this and the functional approach is provable in  $\text{WWKL}_0$ . If there is no suitable Boolean algebra of sets whose completion yields measurable sets, the situation becomes more complicated and the question of whether a similar procedure exists for arbitrary metric spaces still remains open.

## 4.2 Products and Powers in $L_p$ Spaces

Giving a meaningful, viable definition of products and powers of elements of  $L_p(X)$  may seem like an innocuous problem at first, but many difficulties surround it, as we will show below. Even in classical measure theory, a product of integrable functions is not necessarily integrable. It is not at all obvious how to give a definition of this concept that will be natural in second-order arithmetic, and at the same time capture classical properties of products. Such a definition is necessary, as knowing the behavior of  $T(fg)$  for integrable functions  $f$  and  $g$  is necessary for proving the pointwise ergodic theorem. More importantly, products and powers are essential if we hope to emulate classical measure theory as closely as possible. For

example, integrating a function  $f$  over a set  $A$  will require knowing that  $f\chi_A$  is integrable, and to consider  $L_p$  spaces, we will need to understand the behavior of  $f^p$  when  $p$  is a real number.

The goal is to develop as much theory as possible in  $\text{RCA}_0$ , yet this system, for the most part, cannot reason about pointwise properties of functions. Worse yet, a natural definition of products can be sensitive to representations. It is possible to construct a function  $f$  represented with the sequence of simple functions  $\langle f_n \rangle$  such that  $f = 0$ , but  $\langle f_n^2 \rangle$ , though pointwise convergent to 0, does not converge in norm. Such a function can, for example, be defined in the following way:

$$f_n(x) = \begin{cases} 2n^3x & 0 \leq x \leq \frac{1}{2n^2} \\ 2n - 2n^3x & \frac{1}{2n^2} < x \leq \frac{1}{n^2} \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

Each  $f_n$  is nonzero only on an interval of length  $\frac{1}{n^2}$  and peaks at  $x = \frac{1}{2n^2}$ , with  $y = n$ . The measure of each  $f_n$  is  $\frac{1}{2n}$  and a simple computation shows that  $\langle f_n \rangle$  is a Cauchy sequence in  $L_1([0, 1])$  (converging to the zero function), while  $\langle f_n^2 \rangle$  is not.

This means that if one tries to define the product of two functions by the products of their approximations, it can happen that  $f_1 = f_2$  and  $g_1 = g_2$ , but  $f_1g_1 \neq f_2g_2$ . The question that the function defined with (4.1) raises is this: do we want to structure the definition of products so that  $f^2$  exists and coincides with the pointwise product (0 in this case) or do we want it to somehow reflect differences in representations? I have opted for the first solution. After all, in classical measure theory products are defined pointwise.

A measure of how good our characterization of integrable products is will be based on the following considerations. Is it in accord with the classical definition? It must not deem integrable products that classically are not. On the other hand, the class of functions satisfying the definition must not be too small. Ideally, the following should be accomplished: the product should be unique and not depend on representations,  $\text{WWKL}_0$  should be able to show that the product obtained from the definition coincides with the pointwise product when the latter is in  $L_p(X)$ , and  $\text{ACA}_0$  should be able to prove that if  $fg = h$  according to the definition, then  $fg = h$  pointwise a.e.

It makes sense to start from the pointwise product of two functions. Classically, if  $f$ ,  $g$ , and  $h$  are elements of  $L_p(X)$ ,  $h$  is said to be the product of  $f$  and  $g$  if  $h(x) = f(x)g(x)$  a.e. This definition is meaningful in our frame-

work, and it can be proved in  $\text{WWKL}_0$  that the product of two functions, if it exists, is unique and does not depend on representations. The problem is that, though it is capable of determining whether  $h$  is the product of  $f$  and  $g$ , the definition does not tell us how to obtain it. In addition, it uses pointwise properties of functions. Nevertheless, it can be modified to make sense in  $\text{RCA}_0$ .

**Definition 4.2.1** ( $\text{RCA}_0$ ) *If  $f$ ,  $g$  and  $h$  are elements of  $L_p(X)$ , we say that  $h = fg$  and call  $h$  the pointwise product of  $f$  and  $g$  if there exist some representations  $\langle f_n \rangle$ ,  $\langle g_n \rangle$  and  $\langle h_n \rangle$  of  $f$ ,  $g$  and  $h$  respectively, such that  $h(x) = f(x)g(x)$  whenever  $f(x) = \lim_n f_n(x)$ ,  $g(x) = \lim_n g_n(x)$  and  $h(x) = \lim_n h_n(x)$  all exist.*

Although it makes sense, this definition does not seem to be very useful. It guarantees that the pointwise product is independent of representations, but not that it is unique: this fact requires  $\text{WWKL}_0$ . For all we know, there may be no points in the domains of the functions in questions. Because of this, and because this definition provides no insight into the construction of products as elements of  $L_p(X)$ , we consider instead the following, stronger characterization.

**Definition 4.2.2** ( $\text{RCA}_0$ ) *If  $f, g$  and  $h$  are elements of  $L_p(X)$  and if there exist representations  $\langle f_n \rangle$  of  $f$  and  $\langle g_n \rangle$  of  $g$  such that  $\|h - f_n g_n\|_p \rightarrow 0$  with a fixed rate of convergence for some  $h \in L_p(X)$ , then we say that  $h$  is the strong product of  $f$  and  $g$ .*

Notice that if the sequence  $\langle f_n g_n \rangle$  is Cauchy with a fixed rate of convergence, then the strong product  $fg$  exists.

**Proposition 4.2.3** ( $\text{RCA}_0$ ) *If  $fg = h$  in the strong sense, then  $fg = h$  pointwise.*

*Proof.* Assume  $\langle f_n g_n \rangle$  converges to  $h$  in  $L_p$  norm, and let  $x$  be such that  $f(x), g(x)$  and  $h(x)$  are defined. Then, since  $\|h - f_n g_n\|_p \rightarrow 0$  with a fixed rate of convergence, one can construct subsequences  $\langle f'_n \rangle$  and  $\langle g'_n \rangle$  of  $\langle f_n \rangle$  and  $\langle g_n \rangle$  such that  $\|h - f'_n g'_n\|_p \leq 2^{-n}$  for all  $n$ .

Let  $f'$  be the function given by  $\langle f'_n \rangle$ , let  $g'$  be the function given by  $\langle g'_n \rangle$ , and let  $h'$  be the function represented as  $\langle h'_n \rangle$ . Then  $f' = f$ ,  $g' = g$ , and  $h' = h$ , and whenever  $f'(x)$ ,  $g'(x)$ , and  $h'(x)$  are all defined,  $h'(x) = \lim_n f'_n(x)g'_n(x) = (\lim_n f'_n(x))(\lim_n g'_n(x)) = f'(x)g'(x)$ .  $\square$

**Corollary 4.2.4** (WWKL<sub>0</sub>) *If  $fg = h$  in the strong sense, then  $f(x)g(x) = h(x)$  a.e.*

From now on, we are going to take Definition 4.2.2 to be the working definition of products: the remainder of the section will show why this is justified. The strong product does not depend on representations of  $f$  and  $g$ . It is also easy to show that the product, if it exists, is unique, since if  $h_1 = fg$  and  $h_2 = fg$ , then

$$\|h_1 - h_2\|_p \leq \|h_1 - f_n g_n\|_p + \|f_n g_n - h_2\|_p,$$

which approaches 0 with a fixed rate of convergence. Notice that the definition naturally takes care of our counterexample (4.1) above, as for the representation of the zero function  $\langle 0 \rangle$ , its product with itself is also the zero function, as required.

The main result of this section will be Proposition 4.2.7, which states that the strong product of two functions, one of which is essentially bounded, exists. First we show a weaker result.

**Proposition 4.2.5** (RCA<sub>0</sub>) *If  $f, g \in L_{p,\infty}$ , then the strong product  $fg$  exists and is also in  $L_{p,\infty}$ . In fact, for any choice of representations of  $f$  and  $g$  the sequence  $\langle f_n g_n \rangle$  converges, and to the same function.*

*Proof.* Let  $f$  be represented with  $\langle f_n \rangle$ , such that for all  $n$ ,  $|f_n| \leq M_f$  and  $g$  be represented with  $\langle g_n \rangle$ , with  $|g_n| \leq M_g$  for all  $n$ . To show that the sequence  $\langle f_n g_n \rangle$  is convergent, we show it is Cauchy with a fixed rate of convergence. Let  $m < n$ .

$$\begin{aligned} \|f_n g_n - f_m g_m\|_p &= \|f_n g_n - f_n g_m + f_n g_m - f_m g_m\|_p \\ &\leq \|f_n(g_n - g_m)\|_p + \|g_m(f_n - f_m)\|_p \\ &\leq M_f 2^{-m} + M_g 2^{-m} \rightarrow 0. \end{aligned}$$

Pulling out constants outside of the norm is justified, because all functions in question are simple, and measure is monotonic on simple functions.

Showing that the product does not depend on the representations of  $f$  and  $g$  is analogous. Let  $\langle f_n \rangle$  and  $\langle f'_n \rangle$  represent  $f$ , and  $\langle g_n \rangle$  and  $\langle g'_n \rangle$  represent  $g$ . This means that for all  $n$ ,  $\|f_n - f'_n\|_p \leq 2^{-n+1}$  and  $\|g_n - g'_n\|_p \leq 2^{-n+1}$ . By assumption,  $\|h - f_n g_n\|_p \rightarrow 0$  with a fixed rate of convergence. Recall also that all representations of an essentially bounded function have the same bound.

$$\begin{aligned} \|h - f'_n g'_n\|_p &\leq \|h - f_n g_n\|_p + \|f_n g_n - f'_n g_n\|_p + \|f'_n g_n - f'_n g'_n\|_p \\ &\leq \|h - f_n g_n\|_p M_g 2^{-n+1} + M_f 2^{-n+1}, \end{aligned}$$

which implies that  $\|h - f'_n g'_n\|_p \rightarrow 0$  also (with a fixed rate of convergence). It is immediate that  $|fg| \leq M_f M_g$ .  $\square$

In practice, however, it is easier to show that two functions satisfy the following (stronger) condition ( $\star$ ):

Functions  $f$  and  $g$  are elements of  $L_p(X)$ , represented, respectively as  $\langle f_n \rangle$  and  $\langle g_n \rangle$  and there exist a sequence of functions  $\langle h_n \rangle$  as well as a function  $h$  in  $L_p(X)$  such that  $f_n g_m \rightarrow h_n$  when  $m \rightarrow \infty$  and  $h_n \rightarrow h$  when  $n \rightarrow \infty$ .

To be able to use this characterization, we need to show that it implies Definition 4.2.2.

**Proposition 4.2.6 (RCA<sub>0</sub>)** *If two functions  $f$  and  $g$  in  $L_p(X)$  satisfy property ( $\star$ ), then  $fg$  exists in the sense of Definition 4.2.2.*

*Proof.* Assume  $f_n g_m \xrightarrow{m} h_n \xrightarrow{n} h$ . For each  $n$  let  $m_n$  be such that  $\|h_n - f_n g_{m_n}\|_p \leq 2^{-n}$ . Then it is clear that  $\langle f_n \rangle$  represents  $f$  and  $\langle g_{m_n} \rangle$  represents  $g$  and  $\|h - f_n g_{m_n}\|_p \xrightarrow{n} 0$ .  $\square$

If another assumption is added, namely that one of the functions is essentially bounded, then the other direction holds true as well. A sketch of the proof is as follows. Suppose  $\|h - f_n g_n\|_p \rightarrow 0$  for some representations of  $f$  and  $g$ , and that  $f$  is essentially bounded. According to the first fact below, for a fixed  $n$ , there always exists  $h_n$  such that  $\|h_n - f_n g_m\|_p \xrightarrow{m} 0$ . Then

$$\begin{aligned} \|h - h_n\|_p &\leq \|h - f_n g_n\|_p + \|f_n g_n - f_n g_m\|_p + \|f_n g_m - h_n\|_p \\ &\leq \|h - f_n g_n\|_p + M_f 2^{-n} + \|f_n g_m - h_n\|_p, \end{aligned}$$

which, since  $m$  is arbitrary, implies that  $\|h - h_n\|_p \leq \|h - f_n g_n\|_p + M_f 2^{-n}$ . Letting  $n \rightarrow \infty$ , it follows that  $\|h - h_n\|_p \rightarrow 0$ . This still does not conclude the proof, because the conclusion should hold for any representation of  $f$  and  $g$ . Lemma 4.2.7 will show that this is indeed the case.

Observe now some facts about products implied by ( $\star$ ):

*Fact [RCA<sub>0</sub>]:* The sequence of functions  $\langle h_n \rangle$  in the definition above always exists, since  $\|f_n g_m - f_n g_k\|_p \leq M_{f_n} \|g_m - g_k\|_p$ , where  $|f_n| \leq M_{f_n}$ .

*Fact [RCA<sub>0</sub>]:* The definition is independent of the order in which the limits are taken: if  $f_n g_m \xrightarrow{n} h_m \xrightarrow{m} h$  and  $f_n g_m \xrightarrow{m} w_n \xrightarrow{n} w$ , then  $h = w$ . Furthermore, one limit exists if and only if the other does.



This is because we have  $\|h - w\|_p \leq \|h - h_m\|_p + \|h_m - f_n g_m\|_p + \|f_n g_m - w_n\|_p + \|w_n - w\|_p$  for all  $n$  and  $m$ . For every  $k$  it is possible to choose  $m$  and  $n$  large enough such that  $\|h - w\|_p \leq 2^{-k}$ .

The second part is proved by a slight modification of the argument, since for example  $\|h - w_n\|_p \leq \|h - h_m\|_p + \|h_m - f_n g_m\|_p + \|f_n g_m - w_n\|_p$  for all  $m$  (by the previous fact the sequence  $\langle h_n \rangle$  always exists).

We now come to the promised main result of this section, which will enable us to integrate over arbitrary integrable sets. Notice that  $(\star)$  is precisely the characterization needed to prove the claim.

**Proposition 4.2.7** (RCA<sub>0</sub>) *Let  $f$  and  $g$  be in  $L_p(X)$  for some  $p$ . If  $g$  is essentially bounded, then  $fg$  exists as an element of  $L_p(X)$ . Furthermore, all representations of  $f$  and  $g$  yield the same product.*

*Proof.* Let  $f$  be represented with  $\langle f_n \rangle$ , and  $g$  with  $\langle g_n \rangle$ . Assume that  $g$  is essentially bounded with  $|g| \leq M_g$ .

Fix  $n$ . Since  $f_n \in S(X)$ , it is bounded, and there exists  $M_{f_n}$  such that  $|f_n(x)| \leq M_{f_n}$  for all  $x$ . Consider  $\langle f_n g_k \mid k \in \mathbb{N} \rangle$ . Let  $m > k$ .

$$\|f_n g_k - f_n g_m\|_p \leq M_{f_n} 2^{-k}.$$

Therefore,  $\langle f_n g_k \mid k \in \mathbb{N} \rangle$  is a Cauchy sequence that represents  $f_n g$ .

Next show that the sequence  $\langle f_n g \mid n \in \mathbb{N} \rangle$  is strong Cauchy in  $L_p(X)$ . For  $n > m$

$$\begin{aligned} \|f_m g - f_n g\|_p &= \|(f_m - f_n) g\|_p \\ &\leq M_g \|f_m - f_n\|_p \\ &\leq M_g 2^{-m}. \end{aligned}$$

The first inequality would be almost trivially true in WWKL<sub>0</sub> (compare to Proposition 4.4.1). In RCA<sub>0</sub>, however, it requires some work. It follows from these two facts:

$$\|(f_m - f_n) g_k\|_p \leq M_g \|f_m - f_n\|_p,$$

which is true by monotonicity of measure for simple functions, and

$$\|(f_m - f_n) g_k\|_p \rightarrow \|(f_m - f_n) g\|_p. \quad (4.2)$$

That this is enough follows from the more general fact (with a simple proof) that, if  $\langle x_n \rangle$  is a sequence of real numbers that converges to  $x$ , and if  $y$  is such that for all  $n$ ,  $x_n \leq y$ , then  $x \leq y$  as well. It remains to show (4.2).

It is true that  $\|g - g_k\|_p \leq 2^{-k}$ . Furthermore,

$$\begin{aligned} \|(f_m - f_n)g - (f_m - f_n)g_k\|_p &= \|(g - g_k)(f_m - f_n)\|_p \\ &\leq (M_{f_n} + M_{f_m})2^{-k}, \end{aligned}$$

since  $f_n$  and  $f_m$  are simple functions. This means that  $(f_m - f_n)g_k \rightarrow (f_m - f_n)g$  in  $L_p(X)$  for each fixed  $m$  and  $n$  when  $k \rightarrow \infty$ , which also concludes the proof of the above claim.

Consequently, there exists  $h \in L_p(X)$  such that  $\|h - f_n g\|_p \rightarrow 0$  as  $n \rightarrow \infty$ . By property  $(\star)$ ,  $h = fg$ .

To prove the second part, let  $f$  and  $g$  be as above, such that  $f_n g_m \xrightarrow{m} h_n \xrightarrow{n} h$  for some representations of  $f$  and  $g$  (recall that, if one of the functions is essentially bounded, then property  $(\star)$  coincides with the definition of products). We will show that if  $\langle f'_n \rangle$  and  $\langle g'_n \rangle$  are different representations of  $f$  and  $g$ , there exists a sequence of functions  $\langle h'_n \rangle$  in  $L_p(X)$  such that  $f'_n g'_m \xrightarrow{m} h'_n \xrightarrow{n} h$ .

Recall that for each  $n$ ,  $h'_n$  exists. Since each  $f'_n$  is a simple function, for every  $n$  there exists a positive constant  $M_{f'_n}$  such that  $|f'_n| \leq M_{f'_n}$ . Next,

$$\begin{aligned} \|h_n - h'_n\|_p &\leq \|h_n - f_n g_m\|_p + \|f_n g_m - f'_n g_m\|_p \\ &\quad + \|f'_n g_m - f'_n g'_m\|_p + \|f'_n g'_m - h'_n\|_p \\ &\leq \|h_n - f_n g_m\|_p + M_g 2^{-n+1} \\ &\quad + M_{f_n} 2^{-m+1} + \|f'_n g'_m - h'_n\|_p. \end{aligned}$$

Since  $m$  is arbitrary, it follows that  $\|h_n - h'_n\|_p \leq M_g 2^{-n+1}$ . Furthermore,

$$\|h - h'_n\|_p \leq \|h - h_n\|_p + \|h_n - h'_n\|_p \leq \|h - h_n\|_p + M_g 2^{-n+1},$$

which approaches 0 when  $n \rightarrow \infty$  and with a fixed rate of convergence.  $\square$

The following corollary contributes evidence to the claim that essentially bounded functions have particularly nice properties with respect to products.

**Corollary 4.2.8 (WWKL<sub>0</sub>)** *If  $f$  or  $g$  is essentially bounded, and if  $fg = h$  pointwise a.e., then  $fg = h$  in the strong sense.*

*Proof.* Since the strong product in this case exists, and is unique, the two have to coincide.  $\square$

Thus, when  $f$  or  $g$  is essentially bounded, Definition 4.2.2 provides a characterization of the pointwise product that is useful in the absence of WWKL<sub>0</sub>, but provably equivalent to Definition 4.2.1 in the presence of WWKL<sub>0</sub>.

Let us recapitulate the results of this section.

In  $\text{RCA}_0$  we have the following:

- If one of the functions is essentially bounded, the strong product exists, is unique and does not depend on representations.
- The pointwise product, if it exists, does not depend on representations.
- The strong product, if it exists, is unique.
- If  $h$  is the strong product of  $f$  and  $g$ , then it is their pointwise product.

In  $\text{WWKL}_0$  the following hold:

- The strong product, if it exists, does not depend on representations.
- The pointwise product, if it exists, is unique.
- If  $h$  is the strong product of  $f$  and  $g$ , then  $h = fg$  a.e.
- If one of the functions is essentially bounded, and  $h = fg$  a.e, then  $h = fg$  in the strong sense.

The situation at hand is satisfactory: when one of the functions is essentially bounded, we have all the necessary results in  $\text{RCA}_0$ . In general, however, Definition 4.2.1 and Definition 4.2.2 are not equivalent for more general functions; in other words, when neither  $f$  nor  $g$  is essentially bounded, they may have a pointwise product that is not a product in the strong sense.

In light of these difficulties, we cannot safely say that the suggested definition is necessarily the best one. The most important thing is that it works, which it does, but it is not difficult to imagine that there is a more natural, more elegant, more accessible (with respect to  $\text{RCA}_0$ ) way to deal with products of integrable functions.

Defining powers of functions is also tricky. (Before continuing, note that the discussion below makes sense only for nonnegative functions.) Even defining powers of numbers is not trivial. We tend to forget that, while defining  $x^2$  is easy,  $x^{\sqrt{3}}$ , or even  $x^{1/2}$ , requires some thought. All the definitions and necessary properties of power functions are given in Appendix B.

As with products, we can give two definitions. The first is pointwise.

**Definition 4.2.9** *Let  $f \in L_p(X)$  and let  $k \in \mathbb{R}$ . If there exists a representation  $\langle f_n \rangle$  of  $f$  and  $h \in L_p(X)$  such that whenever  $x$  is in the domain of  $h$ , and  $\lim_n f_n(x) = f(x)$  exists, and if  $f_n^k(x) = h(x)$ , we say that  $h$  is the pointwise  $k^{\text{th}}$  power of  $f$ .*

The other alternative is to consider convergence in norm.

**Definition 4.2.10** *Let  $f \in L_p(X)$  and let  $k \in \mathbb{R}$ . If there exists a representation  $\langle f_n \rangle$  of  $f$  such that  $\|h - f_n^k\|_p \rightarrow 0$  for some  $h \in L_p(X)$ , we say that  $h$  is the strong  $k^{\text{th}}$  power of  $f$ .*

The relationships between the two definitions in  $\text{RCA}_0$  and  $\text{WWKL}_0$  are the same as the ones for products.

Definition 4.2.10 is closely related to Definition 4.2.2, but is weaker. Although we would hope that the two would coincide when considering  $f^2$  for some function  $f$ , it is not necessarily the case: if  $f^2$  exists in the sense just defined, then  $f \cdot f$  exists as well. The reverse is still an open question: it is conceivable that two different representations of  $f$  are taken to form the product, whereas for no representation is  $\langle f_n^2 \rangle$  convergent. As before, this issue can be resolved in stronger systems. Also, as usual, if  $f$  is an element of  $L_{p,\infty}$ , the definitions do coincide. Assuming  $\langle f_n \rangle$ ,  $\langle f'_n \rangle$  and  $\langle f''_n \rangle$  are all representations of an essentially bounded function with bound  $M_f$ , and there is some function  $h$  such that  $\|h - f'_n f''_n\|_p \rightarrow 0$ , the conclusion follows from the fact that

$$\begin{aligned} \|h - f_n^2\|_p &\leq \|h - f'_n f''_n\|_p + \|f'_n f''_n - f'_n f_n\|_p + \|f'_n f_n - f_n^2\|_p \\ &\leq \|h - f'_n f''_n\|_p + M_f 2^{-n+1} + M_f 2^{-n+1}. \end{aligned}$$

For this reason, as in the case of products, it would be reasonable to restrict the definition to essentially bounded functions only.

### 4.3 Further Properties of $L_p$ Spaces

The next task is to determine which of the standard properties of  $L_p$  spaces carry over to our framework, and for those that do, whether the same proofs can be used and in which subsystem.

First we prove Hölder's inequality.

**Lemma 4.3.1** ( $\text{RCA}_0$ ) *If  $f \in L_p(X)$  and  $g \in L_q(X)$  (where  $p > 1$ ,  $q > 1$  and  $\frac{1}{p} + \frac{1}{q} = 1$ ) are given with representations  $\langle f_n \rangle$  and  $\langle g_n \rangle$ , then the sequence  $\langle f_n g_n \rangle$  is strong Cauchy in  $L_1(X)$ , therefore  $fg$  is integrable and  $\|fg\|_1 \leq \|f\|_p \|g\|_q$ .*

*Proof.* In the standard proof, integrability of  $fg$  is not established separately, as it is a consequence of Hölder's inequality:  $\int fg < \infty$ , which is enough in the classical setup. We cannot follow this reasoning, since  $fg$  cannot be

integrated without the knowledge that it is integrable in the sense of one of the definitions of products.

We are able to show, however, that for every  $n$

$$\|f_n g_n\|_1 \leq \|f_n\|_p \|g_n\|_q, \quad (4.3)$$

using the standard argument. The inequality  $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$  holds for all nonnegative real numbers  $a$  and  $b$  is provable in  $\text{RCA}_0$ , since a basic theory of differentiation is available in this system, and the proof consists of finding the minimum of the function  $b \mapsto ab - \frac{a^p}{p} + \frac{b^q}{q}$  (see Appendix B for a proof of this inequality). Let  $u$  and  $v$  be any two simple functions. Let  $a = \frac{|u(x)|}{\|u\|_p}$  and  $b = \frac{|v(x)|}{\|v\|_q}$  in the above inequality and obtain

$$\frac{|u(x)|}{\|u\|_p} \cdot \frac{|v(x)|}{\|v\|_q} \leq \frac{1}{p} \frac{|u(x)|^p}{\|u\|_p^p} + \frac{1}{q} \frac{|v(x)|^q}{\|v\|_q^q}.$$

Integrating yields the desired inequality. In particular, when  $u = f_n$  and  $v = g_n$ , (4.3) follows.

Next, show that  $fg$  is integrable. Let  $m < n$ . Since  $f_n - f_m$  and  $g_n - g_m$  are simple, Hölder's inequality for simple functions applies.

$$\begin{aligned} \|f_n g_n - f_m g_m\|_1 &= \|f_n g_n - f_n g_m + f_n g_m - f_m g_m\|_1 \\ &\leq \|f_n(g_n - g_m)\|_1 + \|g_m(f_n - f_m)\|_1 \\ &\leq \|f_n\|_p \|g_n - g_m\|_q + \|g_m\|_q \|f_n - f_m\|_p. \end{aligned}$$

Every Cauchy sequence of real numbers is bounded:  $\|f_n\|_p \leq \|f_1\|_p + 2^{-1} = M_1$  and  $\|g_n\|_1 \leq \|g_1\|_1 + 2^{-1} = M_2$ . Then

$$\|f_n g_n - f_m g_m\|_1 \leq M_1 2^{-m} + M_2 2^{-m},$$

which is a Cauchy sequence with a fixed rate of convergence. The function  $fg$  is integrable, and taking the limit in (4.3) is justified. The Hölder's inequality follows.  $\square$

It is worth pointing out that our definition of  $L_p$  spaces is different from that of most authors. Even in Bishop's work, a function  $f$  is an element of  $L_p(X)$  if and only if  $f$  is measurable and  $|f|^p \in L_1(X)$ . How are the two definitions related? An exact correspondence cannot be established, since we do not have the notion of measurable function. The measurability condition has to be replaced with another, though, as it is easy to construct a sequence of simple functions  $\langle f_n \rangle$  which is not Cauchy, whereas for some  $p$   $\langle f_n^p \rangle$  is (for

example,  $p = 2$ ,  $f_{2n}(x) = 1$  and  $f_{2n+1}(x) = -1$ ). No such difficulty arises when functions considered are nonnegative, so it is reasonable to give the following characterization. Interestingly enough, its proof is not as obvious as one might imagine.

**Lemma 4.3.2 (RCA<sub>0</sub>)** *A function  $f$  is an element of  $L_p(X)$  if and only if  $(f^+)^p$  and  $(f^-)^p$  are elements of  $L_1(X)$ .*

*Proof.* One direction is straightforward. Let  $f = \langle f_n \rangle$  and assume  $(f^\pm)^p \in L_1(X)$ . The inequality  $|x - y|^p \leq |x^p - y^p|$  holds for all  $x, y \in \mathbb{R}$  (proved in Appendix B) and implies that

$$|f_n^\pm - f_m^\pm|^p \leq |(f_n^\pm)^p - (f_m^\pm)^p|$$

for all  $n, m$ . Applying  $\mu$  to the inequality yields

$$\|f_n^\pm - f_m^\pm\|_p^p \leq \|(f_n^\pm)^p - (f_m^\pm)^p\|_1 \leq 2^{-m}$$

when  $m < n$ . Therefore,  $\langle f_n^+ \rangle$  and  $\langle f_n^- \rangle$  are Cauchy in  $L_p(X)$ , meaning that  $f^+ \in L_p(X)$  and  $f^- \in L_p(X)$ ; consequently  $f = f^+ - f^- \in L_p(X)$ .

For the other direction it suffices to show that if  $f$  is nonnegative and  $f \in L_p(X)$ , then  $f^p \in L_1(X)$  (since  $f = f^+ - f^-$ , and  $f^+$  and  $f^-$  are nonnegative).

Based on the inequality  $|x^p - y^p| \leq p|(x - y)(x^{p-1} + y^{p-1})|$ , which is true for all real  $x$  and  $y$  (for proof see Appendix B),

$$|f_n^p - f_m^p| \leq p|(f_n - f_m)(f_n^{p-1} + f_m^{p-1})|$$

for all  $n$  and  $m$ . Assume  $m < n$ . After integrating, and with the aid of Hölder's inequality,

$$\begin{aligned} \|f_n^p - f_m^p\|_1 &\leq p \|f_n - f_m\|_p \|f_n^{p-1} + f_m^{p-1}\|_q \\ &\leq p \|f_n - f_m\|_p (\|f_n^{p-1}\|_q + \|f_m^{p-1}\|_q), \end{aligned}$$

where  $q$  is such that  $\frac{1}{p} + \frac{1}{q} = 1$ .

Because  $f \in L_p(X)$ ,  $\|f_n - f_m\|_p \leq 2^{-m}$ . It remains to estimate  $\|f_n^{p-1}\|_q$  and  $\|f_m^{p-1}\|_q$ . Notice that  $p = (p - 1)q$ .

$$\|f_n^{p-1}\|_q = [\mu(f_n^{(p-1)q})]^{1/q} = [\mu(f_n^p)]^{1/q} = \|f_n\|_p^{p/q}.$$

The sequence  $\langle \|f_n\|_p \rangle$  is Cauchy in  $\mathbb{R}$  and thus bounded; let  $M$  be an upper bound. Then

$$\|f_n^p - f_m^p\|_1 \leq p M^{p/q} 2^{-n+1},$$

and  $f^p$  is an integrable function.  $\square$

Next we can determine the relationship among the  $L_p(X)$  spaces. The proof is adapted from [27].

**Lemma 4.3.3** (RCA<sub>0</sub>) For  $p < q$ ,  $L_q(X) \subseteq L_p(X)$ .

*Note:* The spaces  $L_p(X)$  and  $L_q(X)$  don't really exist as objects. The statement  $L_q(X) \subseteq L_p(X)$  is just an abbreviation for "every function  $f$  that can be represented as a Cauchy sequence in the  $L_q$  norm has an equivalent representation (in the  $L_q$  sense) which is Cauchy in  $L_p(X)$ ."

*Proof.* It is sufficient to show that if  $\langle f_n \rangle \in L_q(X)$ , then  $\langle f_n \rangle \in L_p(X)$ , that is, if  $\|f_m - f_n\|_q \leq 2^{-m}$  for all  $m$  and  $n$  such that  $m < n$ , then  $\|f_m - f_n\|_p \leq 2^{-m}$  as well. It is clearly enough to show that  $\|g\|_p \leq \|g\|_q$  for  $g \in S(X)$ . The claim is derived from the following: for  $x, y \in \mathbb{R}^+$ ,  $y^{q/p} \geq \frac{q}{p} x^{\frac{q-p}{p}} (y - x) + x^{q/p}$ . This inequality is provable in RCA<sub>0</sub> (proved in Appendix B). In particular, let  $f = |g|^p$  ( $f$  is in  $C(X)$ ) and let  $y = f(t)$ :

$$[f(t)]^{q/p} \geq \frac{q}{p} x^{\frac{q-p}{p}} (f(t) - x) + x^{q/p}.$$

Recall that  $\mu(1) = 1$ . Since measure is monotonic on  $C(X)$ , it follows that

$$\mu(f^{q/p}) \geq \frac{q}{p} x^{\frac{q-p}{p}} (\mu(f) - x) + x^{q/p},$$

and since the above inequality is true for all  $x \in \mathbb{R}$ , let  $x = \mu(f)$ , obtaining

$$\mu(f^{q/p}) \geq (\mu(f))^{q/p},$$

which after some algebraic manipulation turns into  $\|g\|_q \geq \|g\|_p$ .  $\square$

The immediate consequence of the above proof is that  $L_2(X) \subseteq L_1(X)$ , a fact needed in the passage from  $L_2$  convergence to  $L_1$  convergence in the ergodic theorem. Furthermore, if  $p < q$  and  $f \in L_q(X)$ , then  $\|f\|_p \leq \|f\|_q$  (proof: let  $f = \langle f_n \rangle$  and, as in the proof of 4.3.3,  $\|f_n\|_q \leq \|f_n\|_p$  for each  $f_n$ ; to obtain the result, take the limit).

Lemma 4.3.3 also implies that every function in  $L_p(X)$ , for any  $p > 1$ , is pointwise defined a.e. (as it is also an element of  $L_1(X)$ ).

## 4.4 Integrable Functions and Sets

In the remainder of the chapter we restrict our attention to the space  $L_1(X)$ .

The proof of the following statements can be found in [31]:

**Proposition 4.4.1** (WWKL<sub>0</sub>) *For  $f, f' \in L_1(X)$  the following properties hold:*

1.  $\|f - f'\| = 0$  if and only if  $f = f'$  a.e.
2. If  $f \leq f'$  a.e, then  $\mu(f) \leq \mu(f')$ .

From now on, “ $f = g$  in  $L_1(X)$ ” and “ $f = g$  a.e. ” will be used interchangeably when the underlying system is WWKL<sub>0</sub> or ACA<sub>0</sub>.

**Lemma 4.4.2** (ACA<sub>0</sub>) *For  $f \in L_1(X)$  and any  $c \in \mathbb{R}$  the sets  $\{x \mid f(x) > c\}$  and  $\{x \mid f(x) \geq c\}$  are integrable.*

*Proof.* Set  $f = \langle f_n \rangle$ , with  $f_n \in C(X)$ . For each  $n$ ,  $\{x \mid f_n(x) > c\}$  is open, and  $f(x) = \lim_n f_n(x)$  for all  $x$  in the domain of  $f$ , a full  $F_\sigma$  set  $F$ . Since characteristic functions of sets that differ by a null set are equal in  $L_1(X)$ , there is no loss of generality in assuming that  $x \in X$  instead of  $x \in F$ .

$$\begin{aligned} \{x \mid f(x) > c\} &= \{x \mid \exists m \exists k \forall n \geq k f_n(x) > c + 2^{-m}\} \\ &= \bigcup_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n \geq k} \{x \mid f_n(x) > c + 2^{-m}\}, \end{aligned}$$

which is integrable (in ACA<sub>0</sub>). The second claim follows from the first, using complements.  $\square$

It is not difficult to show that the previous claim reverses to ACA<sub>0</sub>.

**Theorem 4.4.3** (RCA<sub>0</sub>) *The following are equivalent:*

1. For  $f \in L_1(X)$  and any  $c \in \mathbb{R}$  the set  $\{x \mid f(x) > c\}$  is integrable.
2. (ACA).

*Proof.* That (1)  $\rightarrow$  (2) was proved in Lemma 4.4.2. To prove the other direction, we will use item 2. from Lemma 1.2.3. For that purpose, let a sequence  $\langle a_n \rangle$  of nonnegative numbers is given, such that for all  $n$ ,  $\sum_{i < n} a_i \leq 1$ . We are going to construct a function  $f \in L_1([0, 1])$  such that

$$\mu(\{x \mid f(x) > 0\}) = \sum_n a_n. \quad (4.4)$$



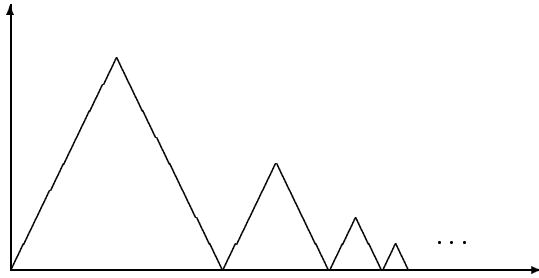
For this purpose, define a Cauchy sequence of simple functions  $\langle f_n \rangle$  such that  $\mu(f_n) = \sum_{i \leq n} \frac{a_i}{2^i}$ . This can be accomplished as follows. First define  $f_1$ :

$$f_1(x) = \begin{cases} \frac{2}{a_1}x & 0 \leq x < \frac{a_1}{2} \\ 1 - \frac{2}{a_1}x & \frac{a_1}{2} \leq x < a_1 \\ 0 & a_1 \leq x. \end{cases}$$

Clearly,  $\mu(f_1) = a_1$ . Similarly, assuming  $f_i$  is defined for  $i < n$ , assuming it satisfies (4.4), define

$$f_n(x) = \begin{cases} f_i(x) & \sum_{k \leq i-1} a_k \leq x < \sum_{k \leq i} a_k, \quad i < n, \\ \frac{2^n}{a_n}x, & \sum_{k \leq n-1} a_k \leq x < \sum_{k \leq n-1} a_k + \frac{a_n}{2}, \\ \frac{2^{n-1}}{a_n} - \frac{2^n}{a_n}x, & \sum_{k \leq n-1} a_k + \frac{a_n}{2} \leq x < \sum_{k \leq n} a_k, \\ 0, & \sum_{k \leq n} a_k \leq x. \end{cases}$$

It is probably easier to understand how  $f_n$  is defined from its graph:



where the length of the side of the  $i^{\text{th}}$  triangle lying on the  $x$ -axis is  $a_i$  and each triangle has area  $\frac{a_i}{2^i}$ , hence  $\mu(f_n) = \sum_{i \leq n} \frac{a_i}{2^i}$ . The function corresponding to this graph is a simple function.

First let us show that  $\langle f_n \rangle$  is a strong Cauchy sequence. Notice that the sequence  $\langle f_n \rangle$  is increasing, so if  $m < n$ ,

$$\begin{aligned} \|f_n - f_m\| &= \mu(f_n - f_m) = \mu(f_n) - \mu(f_m) \\ &= \sum_{k=m+1}^n \frac{a_k}{2^k} \leq \sum_{k=m+1}^n \frac{1}{2^k} \leq \frac{1}{2^{m+1}}, \end{aligned}$$

as required. This means that there is a function  $f \in L_1([0, 1])$  represented with this sequence. Furthermore, the set  $M = \{x \mid f(x) > 0\}$  is precisely the domain of  $f$  with the exception of countably many points of the form

$x = \sum_{i \leq n} a_i$ , where  $f(x) = 0$ , but countable sets don't affect measure. Since  $M$  is integrable,  $\mu(M)$  exists, and it is not hard to show that  $\mu(M) = \sum_n a_n$ , which completes the proof.  $\square$

It is conceivable that, with more care, integrability of  $\{x \mid f(x) > c\}$  could be proved in a weaker system than  $\text{ACA}_0$ , for "enough" values of  $c$ . Bishop developed the entire theory of profiles (see [8]) for this purpose. He defined *accessible* numbers and showed that the sets of the above form are integrable if  $c$  is accessible. Without a doubt, it would be worthwhile to examine Bishop's approach more carefully and see if his ideas can be utilized in our framework, but this falls outside of the scope of this work. Since the proof of the pointwise ergodic theorem requires arithmetic comprehension in a number of places, this is not a great loss.

We also want to integrate over arbitrary integrable sets. Since  $g = f\chi_G$  is integrable by Proposition 4.2.7, it is valid to define  $\int_G f = \mu(f\chi_G)$ .

Similarly, to define a function by cases, for  $f$  to be  $f_k$  on  $M_k$ , where  $f_k \in L_1(X)$  and  $M_k$  a integrable set, define

$$f = \sum_{k=1}^n f_k \chi_{M_k}.$$

Linear combinations of integrable functions are integrable, and since  $f_k \chi_{M_k}$  is integrable for each  $k$ ,  $f$  is integrable.

We now adapt two standard results from measure theory, the first of which will be used in the first proof of the pointwise ergodic theorem, and the second to prove one of its consequences. Both proofs are similar to standard ones, though the second uses the monotone convergence theorem and is postponed until the next section.

**Lemma 4.4.4** ( $\text{ACA}_0$ ) *Let  $\langle f_n \rangle$  be a pointwise convergent sequence of functions in  $L_1(X)$ , which converges in norm to an integrable function  $f$ . Then  $\langle f_n \rangle$  converges to  $f$  pointwise a.e.*

*Proof.* First show that  $\langle f_n \rangle$  has a subsequence that converges pointwise to  $f$ . Since  $\|f - f_n\| \rightarrow 0$ , for each  $k$  let  $f_{n_k}$  be such that  $\|f - f_{n_k}\| \leq 2^{-k}$ .

We are going to show that  $\mu(\{x \mid \lim_k f_{n_k}(x) \neq f(x)\}) = 0$ , which will, by regularity of measure for  $G_\delta$  sets, implies that  $\lim_k f_{n_k}(x) = f(x)$  a.e.

Let  $U = \{x \mid \lim_k f_{n_k}(x) \neq f(x)\}$ . Since

$$\lim_k f_{n_k}(x) \neq f(x) \leftrightarrow \exists m \forall l \exists k \geq l (|f(x) - f_{n_k}(x)| \geq 2^{-m}),$$

the set  $U$  can be represented as

$$U = \bigcup_m \bigcap_l \bigcup_{k \geq l} U_{mk},$$

where  $U_{mk} = \{x \mid |f(x) - f_{n_k}(x)| \geq 2^{-m}\}$ .

Let us evaluate the measure of  $U_{mk}$ . Because  $U_{mk}$  is integrable, we can integrate over this set.

On the one hand,  $\int_{U_{mk}} |f_{n_k} - f| \leq \|f_{n_k} - f\| \leq 2^{-k}$ . On the other,  $\int_{U_{mk}} |f_{n_k} - f| \geq \int_{U_{mk}} 2^{-m} = 2^{-m} \mu(U_{mk})$ . Therefore,  $\mu(U_{mk}) \leq 2^{m-k}$ .

Next,  $\mu(\bigcup_{k \geq l} U_{mk}) \leq \sum_{k \geq l} 2^{m-k} = 2^{m-l+1}$ , which implies

$$\mu(\bigcap_l \bigcup_{k \geq l} U_{mk}) \leq 2^{m-l+1}$$

for all  $l$ , and hence this set has measure 0. The set  $U$ , consequently, is the union of sets of measure 0, so itself has measure 0. As was mentioned earlier, since it is simple enough in structure,  $U$  is contained in a null  $G_\delta$  set, so it can be assumed that  $U$  is a  $G_\delta$  set. Hence, the subsequence  $\langle f_{n_k} \rangle$  converges a.e. to  $f$ .

This concludes the first part of the proof. It remains to show that the entire sequence converges pointwise to  $f$ . By pointwise convergence of  $\langle f_n \rangle$ , for all  $x$  outside of a null  $G_\delta$  set there exists some  $\hat{x}$  such that  $f_n(x) \rightarrow \hat{x}$ . It remains to show that  $\hat{x} = f(x)$  a.e. Since

$$|\hat{x} - f(x)| \leq |\hat{x} - f_{n_k}(x)| + |f_{n_k}(x) - f(x)|,$$

it is easy to see that this is true.  $\square$

Compare this lemma to the result from Section 4.2 that if  $\langle f_n g_n \rangle$  converges strongly to  $h$ , then it converges pointwise product as well. The proof of this fact only needed  $\text{RCA}_0$ , whereas Lemma 4.4.4 is proved in  $\text{ACA}_0$ . There is no contradiction between the two. The functions in Section 4.2 were simple, and the convergence was strong: we have neither of the two in the previous lemma.

With additional assumptions, the reverse also holds, i.e. pointwise convergence can imply convergence in norm.

**Lemma 4.4.5** ( $\text{ACA}_0$ ) *If  $\langle f_n \rangle$  is a sequence of integrable functions that converges pointwise to an integrable function  $f$ , and if in addition  $\|f_n\| \leq \|f\|$  (or  $\|f_n\| \rightarrow \|f\|$ ) then  $f_n \rightarrow f$  in  $L_1(X)$ .*

A proof is given in the next section.

## 4.5 The Monotone Convergence Theorem

The monotone convergence theorem will be used in a number of places in the next chapter, directly, or through one of its consequences. It is also a convenient tool for proving integrability of certain functions. In [31], Yu showed that the dominated convergence theorem is equivalent to  $(ACA)$ , while a weaker version of the monotone convergence theorem is equivalent to  $(WWKL)$  over  $RCA_0$ . In the latter, the limit is a known integrable function. We have little use for that theorem, since in general the sequence of functions will converge at almost every point, but the limit function will not be known in advance. Since the proof of the main theorem is formalized in  $ACA_0$ , it is not a great loss that this stronger version of the monotone convergence theorem, presented below, is also stated and proved in  $ACA_0$ .

**Theorem 4.5.1** ( $ACA_0$ ) *Assume  $\langle f_n \rangle$  is a monotonic sequence of functions in  $L_1(X)$  with bounded measure. Then there is  $f \in L_1(X)$  such that  $\|f_n - f\| \rightarrow 0$ , and  $\mu(f_n) \rightarrow \mu(f)$ .*

*Proof.* Without loss of generality, let  $\langle f_n \rangle$  be increasing, and  $\mu(f_n) \leq M$  for all  $n$ . Since  $\langle f_n \rangle$  is increasing, the sequence  $\langle \mu(f_n) \rangle$  is an increasing sequence of real numbers (Proposition 4.4.1) and as it is bounded, it is convergent and therefore Cauchy so

$$\forall \varepsilon \exists N \forall n > m > N (|\mu(f_n) - \mu(f_m)| < \varepsilon)$$

and

$$|\mu(f_n) - \mu(f_m)| = \mu(f_n) - \mu(f_m) = \mu(f_n - f_m) = \mu(|f_n - f_m|),$$

which means that

$$\forall \varepsilon \exists N \forall n > m > N (\|f_n - f_m\| \leq \varepsilon).$$

The sequence  $\langle f_n \rangle$  is Cauchy, hence convergent in  $ACA_0$ . Because  $L_1(X)$  is complete, the sequence converges to an integrable function  $f$  such that  $\mu(f_n) \rightarrow \mu(f)$ .  $\square$

The proof of the theorem does not change if the sequence  $\langle f_n \rangle$  is assumed to be decreasing and bounded below instead. This would not be true were the space  $X$  not of finite measure.

In both this reversal and that of the pointwise ergodic theorem, it will be necessary to show that a function of the form  $\sum_k c_k \chi_{I_k}$  (where  $I_k$  are disjoint, half-open intervals that cover  $[0, 1]$ ) is integrable. In both cases, the following situation arises:

**Lemma 4.5.2 (RCA<sub>0</sub>)** Let  $0 = a_0 < a_1 < \dots$ , define  $I_k = [a_k, a_{k+1})$  and  $\cup_k I_k = [0, 1]$ . Then  $\chi_{I_k} \in L_1([0, 1])$  and if  $\langle c_k \rangle$  is a sequence of rational numbers with  $|c_{k+1} - c_k| \leq M$  for all  $n$  and some constant  $M$ , then  $\sum_n c_k \chi_{I_k} \in L_1(x)$ .

*Note:* In examples that we will encounter,  $a_k = 1 - 2^{-k}$ , or the sequence  $\langle a_k \rangle$  is finite.

*Proof.* It is necessary to find a Cauchy sequence (with a fixed rate of convergence)  $\langle g_n \rangle$  of functions in  $C(X)$  such that  $\lim_n g_n = \sum_k c_k \chi_{I_k}$ . Let  $l_k = a_{k+1} - a_k$ . Fix  $n$  and define

$$g_n(x) = c_k, \quad \text{for } a_k + l_k \cdot 2^{-k-n} \leq x \leq a_{k+1} - l_k \cdot 2^{-k-n},$$

(except when  $k = 0$ , in which case the condition is  $0 \leq x \leq a_1 - l_0 \cdot 2^{-n}$ ) and make it linear otherwise. Let  $m < n$ . Then  $g_n$  and  $g_m$  differ only on the open interval

$$(a_{k+1} - l_k \cdot 2^{-k-m}, a_{k+1} + l_{k+1} \cdot 2^{-k-1-m})$$

for each  $k$ . Each of these intervals has length  $l_{k+1} \cdot 2^{-k-1-m} + l_k \cdot 2^{-k-m}$ , which is less than  $2^{-k-m-1}$ , and  $|g_n - g_m| \leq M$  on each, hence

$$\mu(|g_n - g_m|) \leq \sum_k M 2^{-k-m-1} = M 2^{-m+2}.$$

The sequence  $\langle g_n \rangle$  therefore represents an integrable function: the function  $\sum_k c_k \chi_{I_k}$ .  $\square$

The reversal can now be proved.

**Theorem 4.5.3 (RCA<sub>0</sub>)** *The following are equivalent:*

1. (ACA).
2. Monotone convergence theorem for an arbitrary measure on a complete separable metric space.
3. Monotone convergence theorem for the Lebesgue measure on  $[0, 1]$ .

*Proof.* (1)  $\rightarrow$  (2) was proved above and (2)  $\rightarrow$  (3) is immediate. It remains to show (3)  $\rightarrow$  (1). Let  $\langle a_n \rangle$  be a monotonic bounded sequence of real numbers. We may assume  $0 \leq a_1 \leq a_2 \leq \dots \leq 1$ . The goal is to show that this sequence is convergent, which will in turn imply arithmetic comprehension

(by Lemma 1.2.3). We will construct an increasing, pointwise convergent sequence of functions  $\langle f_n \rangle$  in  $L_1[0, 1]$ , such that  $\mu(f_n) = a_n$ . Assume the monotone convergence theorem holds. One of the conclusions of the theorem was that  $\mu(f_n)$  converges. This implies that  $\langle a_n \rangle$  is convergent as well.

It remains to construct the desired sequence of functions.

$$f_n(x) = \begin{cases} a_1 & x \leq \frac{1}{2} \\ 2a_2 - a_1 & \frac{1}{2} < x \leq \frac{3}{4} \\ \dots & \dots \\ 2^{n-1}a_n - 2^{n-2}a_{n-1} - \dots - a_1 & \frac{2^{n-1}-1}{2^{n-1}} < x \leq 1. \end{cases}$$

For each fixed  $n$ ,  $f_n$  is a step function that jumps at every  $\frac{2^k-1}{2^k}$  for  $1 \leq k \leq n-1$ , with step size bounded by  $2^{n-1}$ . According to Lemma 4.5.2,  $f_n \in L_1([0, 1])$  for each  $n$ . Furthermore,  $\mu(f_n) = a_n$ , the sequence is increasing and converges everywhere except possibly at 1, and is therefore as required. This completes the proof.  $\square$

The monotone convergence theorem helps us establish integrability of suprema and infima of sequences of integrable functions in the case their norms are uniformly bounded (or if the functions are essentially bounded with the same bound). This is really the only case of interest for the pointwise ergodic theorem, because the sequence under consideration will be  $\langle S_n f \rangle$ , where, since the transformation  $T$  preserves norm,  $\|S_n f\| \leq \|f\|$  for all  $n$ .

By the supremum of a sequence of functions we mean the  $L_1$  limit of the sequence  $g_n = \max\{f_1, \dots, f_n\}$  (the definitions of  $\inf_n f_n$ ,  $\liminf_n f_n$  and  $\limsup_n f_n$  are analogous). This definition is not to be confused with the pointwise definition: in  $\text{ACA}_0$  for a.e.  $x$ ,  $\sup_n f_n(X)$  exists, but it is not obvious that these pointwise values define an integrable function. Unless otherwise specified, all limits, suprema or infima, and all equalities are meant in the  $L_1$  sense.

**Proposition 4.5.4** ( $\text{ACA}_0$ ) *If  $\langle f_n \rangle$  is a sequence of functions in  $L_1$  such that  $|\mu(f_n)| \leq M$ , then  $\sup_n f_n$ ,  $\inf_n f_n$ ,  $\limsup f_n$  as well as  $\liminf f_n$  are all integrable as well.*

*Proof.* Consider only  $\sup_n f_n$  and  $\limsup_n f_n$  since the argument for  $\inf_n f_n$  and  $\liminf_n f_n$  is analogous. Define  $g_n = \max\{f_1, \dots, f_n\}$ . The sequence  $\langle g_n \rangle$  is increasing and  $\|g_n\| \leq M$ . Theorem 4.5.1 applies:  $\lim_n g_n = \sup f_n$  is an integrable function.

The argument for  $\limsup_n f_n$  is similar. Define  $h_{n,k} = \max_{n \leq m \leq k} f_m$ . Then  $h_{n,k} \in L_1(X)$  and  $\|h_{n,k}\| \leq M$  for all  $k, n$  and, for a fixed  $n$ ,  $\langle h_{n,k} \mid k \in \mathbb{N} \rangle$

$\mathbb{N}$ ) is increasing. By the monotone convergence theorem, for all  $n$ ,  $\lim_k h_{n,k}$  exists and is in  $L_1(X)$ . Set  $h_n = \lim_k h_{n,k} = \sup_{k \geq n} f_k$ , then  $h_n \in L_1(X)$  for all  $n$ .

Applying a similar argument to the sequence  $h_n$  (decreasing sequence of functions on  $L_1(X)$ , bounded in measure), it follows that  $\lim_n h_n$  exists and is in  $L_1(X)$ . But this limit is  $\limsup_n f_n$ , hence  $\limsup_n f_n \in L_1(X)$ , as required.  $\square$

Proposition 4.5.4 and Lemma 4.4.4, imply (in  $\text{ACA}_0$ ) that if a sequence  $\langle f_n \rangle$  converges pointwise to an integrable function  $f$ , then  $\liminf_n f_n = f$ .

**Corollary 4.5.5 (ACA<sub>0</sub>) (Fatou's Lemma)** *Let  $\langle f_n \rangle$  be a sequence of non-negative integrable functions with  $\mu(f_n) \leq M$  for all  $n$ . Then*

$$\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n).$$

*Proof.* Let  $g_n = \inf_{k \geq n} f_k$ . Then by Proposition 4.5.4,  $g_n \in L_1(X)$  for each  $n$ , so  $\langle g_n \rangle$  is an increasing sequence of integrable functions such that  $\mu(g_n) \leq \mu(f_n) \leq M$  for all  $n$ . Since

$$\forall n (\mu(g_n) \leq \mu(f_n)) \rightarrow \liminf_n \mu(g_n) \leq \liminf_n \mu(f_n)$$

and  $\liminf_n \mu(g_n) = \lim_n \mu(g_n) = \mu(\liminf_n f_n)$ , it follows that

$$\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n)$$

as required.  $\square$

Now we can complete the proof of Lemma 4.4.5, promised in the last section.

*Proof.* First observe that  $|f_n| + |f| - |f - f_n| \rightarrow 2|f|$  pointwise. According to the comment after Proposition 4.5.4 on page 66,  $\liminf |f_n| + |f| - |f - f_n| = 2|f|$  and by Fatou's Lemma

$$\begin{aligned} 2\mu(|f|) &= \mu(\liminf_n |f_n| + |f| - |f - f_n|) \leq \liminf_n \mu(|f_n| + |f| - |f - f_n|) \\ &\leq \mu(|f|) + \lim_n \mu(|f_n|) + \liminf_n (-\mu(|f - f_n|)) \\ &= 2\mu(|f|) - \limsup_n \mu(|f - f_n|). \end{aligned}$$

The last inequality yields  $0 \leq \limsup_n \mu(|f - f_n|) \leq 0$ , and it is easy to show that this implies  $\lim_n \mu(|f - f_n|) = 0$ .  $\square$





## Chapter 5

# The Pointwise Ergodic Theorem

The classical statement of the pointwise ergodic theorem is as follows:

If  $T$  is a measure preserving transformation on a metric space  $X$  and  $f$  a function in  $L_1(X)$ , then there exists a function  $\hat{f} \in L_1(X)$  such that the average  $S_n f(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$  converges to  $\hat{f}(x) \in L_1(X)$  a.e. Furthermore, the limit function  $\hat{f}$  is invariant, i.e.  $\hat{f} \circ T = \hat{f}$ .

Before proving the theorem, we need to translate it into our framework. The first thing to consider is the definition of  $T$ . Because there is no such thing as an arbitrary, pointwise defined function from the point of view of second-order arithmetic,  $T$  cannot be defined on the points of the space. Instead, it will be represented as an operator on functions, following the same reasoning as for the mean ergodic theorem. More precisely, instead of the operator on  $X$ , we will consider the operator induced on  $L_1(X)$ . More information on the relationship between these representations can be found in Halmos [16]. The appropriate definition of  $T$  will be discussed fully in Section 5.1.

Three proofs of the pointwise ergodic theorem are presented below. The first is adapted partly from Spitters [27] and partly from Billingsley [3], the second is based on a proof by Katznelson and Weiss [19], which in turn is derived from a proof by Kamae [18] that used nonstandard analysis, while the third mostly follows Billingsley [3]. The first two proofs are in some sense more natural than the third one, though it is more commonly found in textbooks. This is because the first two adopt a functional approach

and avoid the use of invariant sets, whose treatment in our framework is problematic. Nevertheless, as Section 5.6 will show, these difficulties can be overcome. Each proof is interesting in its own right and provides insights into the behavior of measure preserving transformations on the space of integrable functions over a compact metric space.

## 5.1 Motivation

In standard mathematical practice, if  $(X, \mathcal{F}, \mu)$  is a measure space, a measure preserving transformation (m.p.t.)  $T$  is defined as a mapping of  $X$  to itself such that:

- The inverse image of each measurable set is measurable, and
- for each  $A \in \mathcal{F}$

$$\mu(A) = \mu(T^{-1}A),$$

or, equivalently,

$$\mu(f) = \mu(f \circ T)$$

for each  $f \in L_1(X)$ .

When defining a measure preserving transformation within weak subsystems of second-order arithmetic, we cannot do so pointwise. Even if  $T$  is defined on the dense subset  $A$  of  $X$ , as it is not linear, there is no way of extending it to the entire space. Instead, there are two choices before us: to define  $T$  on functions, or on sets. For a number of reasons, the functional approach is preferable. In this case,  $T$  is first defined on simple functions, then extended in the usual way, and we will see later that  $\text{WWKL}_0$  (and sometimes even  $\text{RCA}_0$ ) will suffice to capture the properties that the classical version of the transformation  $T$  has. On the other hand, we saw that there is presently no satisfactory way to formalize integrable sets, other than via characteristic functions. Even when  $X$  is  $[0, 1]$  or  $[0, 1]^n$ , where the characterization of integrable sets is relatively simple, there is no merit in defining  $T$  on sets. It does not make showing that a given set is invariant any easier; besides, the pointwise ergodic theorem is ultimately a statement about points and functions and not about sets. In addition, this method is in tune with the treatment of measure theory by Bishop, and by Simpson and his students. Finally, it matches the approach taken in the first part of the thesis.

To motivate the definition of  $T$  in the context of second-order arithmetic, let us employ briefly classical mathematical reasoning. Assume a

transformation  $T$  is given, measure preserving in the sense specified above. Consider now the induced transformation  $\hat{T}$ , as we did earlier, which is defined as  $\hat{T}f(x) = f(Tx)$  for all  $x$  for which  $Tx$  is in the domain of  $f$ . Then  $\hat{T}$  has the following properties:

- $\hat{T}$  sends  $L_1(X)$  to itself.
- $\hat{T}$  is linear.
- $\hat{T}$  is measure preserving:  $\mu(\hat{T}f) = \mu(f)$  for all  $f \in L_1(X)$ .
- $\hat{T}$  is norm preserving (an isometry):  $\|\hat{T}f\| = \|f\|$  for all  $f \in L_1(X)$ .
- $\hat{T}$  is multiplicative:  $\hat{T}(fg) = \hat{T}f \cdot \hat{T}g$  a.e.
- $\hat{T}$  is nonnegative: if  $f \geq 0$  a.e. then  $\hat{T}f \geq 0$  a.e.
- If  $|f| \leq M$  a.e. then  $|\hat{T}f| \leq M$  a.e. ( $T$  preserves  $L_{1,\infty}$  norm).

*Note:* Such a transformation will be called  $T$  and not  $\hat{T}$  for the remainder of the text. It will be clear from the context that it is an operator on  $L_1(X)$  and not on  $X$ .

The question presents itself of which of these properties characterize the transformation. We will show in the next section that the properties of being norm preserving and measure preserving are equivalent in  $\text{WWKL}_0$  and that nonnegativity is a consequence of linearity and multiplicativity. In proofs of statements that require at least  $\text{WWKL}_0$ , we are going to freely switch between the two notions, and assume  $T$  is an isometry in claims that are proved in  $\text{RCA}_0$ . The only definition given below is that of an isometry, with obvious modifications for a measure preserving transformation. In fact, in the pointwise ergodic theorem, it suffices that  $T$  be nonexpansive with respect to norm, as this alteration does not affect any of the proofs.

Although the main theorem is concerned only with the space  $L_1(X)$ , other  $L_p$  spaces will also be of interest. For this reason, consider first a more general definition. It corresponds exactly to Definition 2.1.6.

**Definition 5.1.1** *An isometry on  $L_p(X)$  is a function  $T : S(X) \rightarrow L_p(X)$  such that*

1.  $T(q_1f_1 + q_2f_2) = q_1Tf_1 + q_2Tf_2$  for  $q_1, q_2 \in \mathbb{Q}$  and  $f_1, f_2 \in S(X)$ ,
2.  $\|Tf\|_p = \|f\|_p$  for every  $f \in S(X)$ .

Then, for  $f = \langle f_n \rangle \in L_p(X)$ , the sequence  $\langle Tf_n \rangle$  is Cauchy in the space because  $T$  is an isometry, and since  $L_p(X)$  is complete,  $Tf$  is well-defined and  $\|Tf\|_p = \|f\|_p$  for all  $f \in L_p(X)$ .

(The alternative is for  $T$  to be nonexpansive, and to change clause 2. to  $\|Tf\|_p \leq \|f\|_p$ . All proofs remain the same with this modification, so assume for simplicity from now on that  $T$  is an isometry.) A function is *invariant* if  $Tf = f$ .

We will mainly be concerned with the case  $p = 1$ .

How does one define a multiplicative transformation? This question is related to that of products of  $L_p$  functions. It seems logical to characterize multiplicativity only on simple functions, and as long as  $fg \in L_p(X)$  it should follow in  $\text{RCA}_0$  that  $T(fg) = Tf \cdot Tg$ . This is not necessarily the case. In fact, as with products, essential boundedness needs to be employed. To ensure that  $Tf \cdot Tg$  is defined, we need to assume that  $T$  takes essentially bounded functions to essentially bounded functions ( $T$  preserves  $L_{p,\infty}$  norm). This is a reasonable assumption, true classically of measure preserving transformations. In practical terms, it guarantees that simple (and test) functions behave well with respect to the transformation.

**Definition 5.1.2** ( $\text{RCA}_0$ ) *A transformation  $T : L_p(X) \rightarrow L_p(X)$  is said to be multiplicative if the following two conditions are satisfied:*

1. *If  $f$  is a simple function with  $|f| \leq M$ , then  $|Tf| \leq M$  also.*
2. *For all  $f, g \in S(X)$ ,  $T(fg) = Tf \cdot Tg$  in  $L_p(X)$ .*

There is no need to demand in 2. that  $Tf \cdot Tg$  be in  $L_p(X)$ , as this is provided by the first condition.

In the preceding definition, product are meant in the sense of Definition 4.2.2, that is, strong products. However, for essentially bounded functions, the product exists in both senses and is the same, so it would be possible to replace the second condition in the definition of multiplicativity with the requirement that  $Tfg(x) = Tf(x)Tg(x)$  for all  $x$  in the domain of  $T(fg)$ ,  $Tf$  and  $Tg$ . The strong definition proves to be more useful.

It can easily be shown that this definition implies that  $T(fg) = Tf \cdot Tg$  when both functions are in  $L_{p,\infty}$ . In particular,  $T$  is multiplicative on  $C(X)$ .

**Proposition 5.1.3** ( $\text{RCA}_0$ ) *If  $\langle f_n \rangle$  and  $\langle g_n \rangle$  are representations of  $f, g \in L_{p,\infty}(X)$ , then  $T(fg) = Tf \cdot Tg$ .*

*Proof.* Assume  $|f| \leq M_f$  and  $|g| \leq M_g$ ; then also  $|Tf| \leq M_f$  and  $|Tg| \leq M_g$ . Notice that  $Tf \cdot Tg$  exists, as do  $Tf_n \cdot Tg$ ,  $T(f - f_n) \cdot Tg$  and  $Tf \cdot T(g - g_m)$

(all functions are essentially bounded) and recall that  $T(f_n g_m) = T f_n \cdot T g_m$  for all  $m$  and  $n$ . Using the triangle inequality,

$$\begin{aligned}
\|T(fg) - Tf \cdot Tg\|_p &\leq \|T(fg) - T(f_n g)\|_p + \|T(f_n g) - T(f_n g_m)\|_p \\
&\quad + \|T f_n \cdot T g_m - T f_n \cdot T g\|_p + \|T f_n \cdot T g - Tf \cdot Tg\|_p \\
&= \|T((f - f_n)g)\|_p + \|T(f_n(g - g_m))\|_p \\
&\quad + \|T f_n T(g_m - g)\|_p + \|T(f_n - f)Tg\|_p \\
&\leq M_g 2^{-n} + M_f 2^{-m} + M_f 2^{-m} + M_g 2^{-n} \\
&= 2M_f 2^{-m} + 2M_g 2^{-n},
\end{aligned}$$

based on the  $(L_1$  and  $L_{1,\infty})$  norm preserving property of  $T$ .

Since the inequality holds for all values of  $m$  and  $n$ ,  $\|T(fg) - Tf \cdot Tg\| = 0$ , as required.  $\square$

It is still unclear if this type of argument can be extended to a more general class of functions.

A seemingly stronger property will be required when considering  $L_p(X)$  for  $p > 1$ , a more general multiplicativity principle:

$$T(f^p) = (Tf)^p,$$

whenever  $f \in S(X)$  and is nonnegative. This property follows from multiplicativity when  $p$  is natural or rational. When  $p \in \mathbb{N}$ , it can be proved by  $\Sigma_1$  induction (the statement “ $T$  is multiplicative on simple functions” is  $\Pi_1$ ) and holds not only for simple, but also for test functions. Since  $f^{\frac{1}{n}} \in C(X)$  if  $f \in S(X)$ , and  $Tf = T([f^{\frac{1}{n}}]^n) = [T(f^{\frac{1}{n}})]^n$ ,

$$(Tf)^{\frac{1}{n}} = T(f^{\frac{1}{n}}),$$

so general multiplicativity holds for all rational numbers. Finally, if  $p$  is an arbitrary real number,  $f^p = \lim_n f^{p_n}$ , where  $p = \langle p_n \rangle$  (see Appendix B for a justification of this fact) and

$$T(f^p) = T(\lim_n f^{p_n}) = \lim_n T(f^{p_n}) = \lim_n (Tf)^{p_n} = (Tf)^p.$$

The second equality is true since  $T$ , as a continuous operator, commutes with limits.

**Definition 5.1.4 (RCA<sub>0</sub>)** *An operator  $T : L_p(X) \rightarrow L_p(X)$  is nonnegative if whenever  $f$  is a nonnegative simple function,  $Tf$  is nonnegative as well.*

We proceed to show that if  $T$  is nonnegative on simple functions, it will preserve nonnegativity of arbitrary functions as well. This fact is almost immediate, with our definition of nonnegativity.

To see this, let  $f$  be a nonnegative integrable function, represented as a sequence of simple functions  $\langle f_n \rangle$ . It was remarked earlier that it is safe to assume that each  $f_n$  is nonnegative. By assumption,  $|Tf_n| = Tf_n$  for all  $n$  and

$$\begin{aligned} \|(|Tf| - Tf)\|_p &\leq \|(|Tf| - |Tf_n|)\|_p + \|(|Tf_n| - Tf_n)\|_p + \|Tf_n - Tf\|_p \\ &\leq 2^{-n} + 0 + 2^{-n} \end{aligned}$$

for all  $n$  because  $T$  is norm preserving, so

$$\|Tf_n - Tf\|_p = \|T(f_n - f)\|_p = \|f_n - f\|_p \leq 2^{-n},$$

and by the triangle inequality,  $\|(|Tf| - |Tf_n|)\|_p \leq \|Tf - Tf_n\|_p$ . Because  $\|(|Tf| - Tf)\|_p \leq 2^{-n}$  for all  $n$ ,  $Tf$  is nonnegative.

We showed earlier in the case of Hilbert spaces that for all  $n$ ,  $T^n$  is continuous, linear and norm preserving. The proof in the case of  $T : L_1(X) \rightarrow L_1(X)$  is no different. Note also that the formula stating that  $T^n$  is multiplicative is also  $\Pi_1$ , so multiplicativity of  $T^n$  is provable using Lemma 1.2.1.

This may be a good place to examine the behavior of  $T$ , as defined in the previous section, on sets. For example, can we prove that  $T$  takes sets to sets? Fortunately, this is the case whenever  $T$  is linear and multiplicative. (If  $T$  is an isometry, then it preserves measures of sets, too.) Let  $A$  be an integrable set, i.e. let  $\chi_A$  be a characteristic function. Since  $\chi_A$  is essentially bounded, multiplicativity applies, and since  $\chi_A^2 = \chi_A$ ,  $T(\chi_A) = T(\chi_A^2) = (T\chi_A)^2$ . In addition,  $T\chi_A$  is essentially bounded by 1 so it is a characteristic function. In standard mathematical practice, with the pointwise definition of  $T$ , if  $\hat{T}$  is the transformation induced by  $T$ , then  $\hat{T}\chi_A = \chi_{T^{-1}A}$ . It may be tempting to think of  $T\chi_A$  in those terms, but we have no way of determining the action of  $T$  on sets, and therefore cannot make this claim.

Recall that the following identities hold in  $\text{RCA}_0$ :

$$\chi_{A \cap B} = \chi_A \cdot \chi_B,$$

$$\chi_{A \cup B} = \chi_A + \chi_B - \chi_A \cdot \chi_B,$$

$$\chi_{A^c} = 1 - \chi_A.$$

After applying  $T$ , they become:

$$T(\chi_{A \cap B}) = T\chi_A \cdot T\chi_B,$$

$$T(\chi_{A \cup B}) = T\chi_A + T\chi_B - T\chi_A \cdot T\chi_B,$$

$$T(\chi_{A^c}) = T(1) - T\chi_A = 1 - T\chi_A.$$

It remains to justify the last identity, i.e. to show that  $T(1) = 1$  (the constant 1 function). If one is careful, this fact can be shown in  $\text{RCA}_0$ , assuming  $T$  is measure preserving. Since 1 is a bounded function,  $T(1)$  is also bounded, and multiplicativity applies.

First, observe that  $T(1) = T(1 \cdot 1) = [T(1)]^2$  and that consequently,

$$(1 - T(1))^2 = 1 - 2T(1) + [T(1)]^2 = 1 - T(1).$$

The multiplication above was legitimate, as  $1 - T(1)$  is a bounded function. We can conclude that  $|1 - T(1)| = 1 - T(1)$  and

$$\|1 - T(1)\|_1 = \mu(1 - T(1)) = \mu(1) - \mu(T(1)) = 0,$$

or  $T(1) = 1$  in  $L_1(X)$ .

The above identities can also be interpreted in the following way. If  $M$  and  $N$  are sets such that  $T\chi_A = \chi_M$  and  $T\chi_B = \chi_N$ , then

$$T(\chi_{A \cap B}) = \chi_{M \cap N},$$

$$T(\chi_{A \cup B}) = \chi_{M \cup N},$$

$$T(\chi_{A^c}) = \chi_{M^c}.$$

Another useful property holds of  $T$ . It seems that pointwise properties of functions cannot be avoided in proving it.

**Proposition 5.1.5** ( $\text{WWKL}_0$ ) *Let  $T : L_1(X) \rightarrow L_1(X)$  be a linear, multiplicative transformation. Then  $T(f^+) = (Tf)^+$  and  $T(f^-) = (Tf)^-$  in  $L_1(X)$ .*

*Proof.* Let  $f$  be represented as  $\langle f_n \rangle$ , where each  $f_n$  is a simple function. The goal is to show that  $\|T(f^\pm) - (Tf)^\pm\|_p = 0$ , where  $\pm$  stands for  $+$  or  $-$ . Notice that for all  $n$ ,  $Tf_n = T(f_n^+) - T(f_n^-) = (Tf_n)^+ - (Tf_n)^-$  and all functions involved are defined a.e, are nonnegative on the domain, and at most one of the terms on each side of the identity is nonzero whenever they are all defined. All these facts immediately imply the claim. For example, if  $Tf_n(x) \geq 0$ , then  $T(f_n^+)(x) \geq 0$ , and  $(Tf_n)^+(x) \geq 0$ , while  $T(f_n^-)(x) = (Tf)^-(x) = 0$ , therefore  $T(f_n^+)(x) = (Tf_n)^+(x)$ . The case  $Tf_n(x) \leq 0$  is analogous. In  $\text{WWKL}_0$  this means that  $\|T(f_n^\pm) - (Tf_n)^\pm\| = 0$  for all  $n$ , which in turn implies the claim.  $\square$

We can now without ambiguity write  $Tf^+$  or  $Tf^-$ . This fact also implies that  $T|f| = T(f^+ + f^-) = (Tf)^+ + (Tf)^- = |Tf|$ , from which the following proposition follows:

**Proposition 5.1.6** (WWKL<sub>0</sub>) *Every measure preserving transformation is nonnegative.*

*Proof.* Let  $|f| = f$ . Then  $Tf = T|f| = |Tf|$ . □

As promised in the previous section, we will show that a linear transformation  $T : L_1(X) \rightarrow L_1(X)$  is measure preserving if and only if it is an isometry.

**Proposition 5.1.7** (WWKL<sub>0</sub>) *A linear transformation  $T$  on  $L_1(X)$  is measure preserving if and only if it is norm preserving.*

*Proof.* If  $T$  is measure preserving, then  $\mu(Tf^+) = \mu(f^+)$  and  $\mu(Tf^-) = \mu(f^-)$ , and

$$\begin{aligned} \mu(|f|) &= \mu(f^+ + f^-) = \mu(f^+) + \mu(f^-) \\ &= \mu(Tf^+) + \mu(Tf^-) = \mu(|Tf|). \end{aligned}$$

For the other direction, let  $T$  be an isometry. It should be clear that  $|f^\pm| = f^\pm$ . Similarly, because  $Tf^\pm$  stands for  $(Tf)^\pm$ ,  $|Tf^\pm| = Tf^\pm$ . Therefore:

$$\begin{aligned} \mu(f) &= \mu(f^+) - \mu(f^-) = \mu(|f^+|) - \mu(|f^-|) \\ &= \mu(|Tf^+|) - \mu(|Tf^-|) \\ &= \mu(Tf^+) - \mu(Tf^-) \\ &= \mu(Tf). \end{aligned}$$

□

Due to the two propositions just proved, from now on we can always assume that  $T$  is a multiplicative isometry when working in theories extending WWKL<sub>0</sub>.

## 5.2 Convergence in Norm

For  $T$  defined as in Section 5.1, for  $n \geq 1$ , let

$$S_n f = \frac{1}{n} \sum_{k=0}^{n-1} T^k f.$$



If  $T$  is an isometry, then  $\|S_n f\| \leq \|f\|$  for all  $n$ . In fact, if  $T$  is measure preserving (recall the two are equivalent in  $\text{WWKL}_0$ ),  $\|S_n f\| = \|f\|$ , though the first property is sufficient for our purposes.

Recall that Chapter 3 established that the mean ergodic theorem is equivalent to  $(ACA)$ . Since  $L_2(X)$  is a real Hilbert space, arithmetic comprehension implies convergence of  $S_n f$  in  $L_2$  norm.

The next step in the proof of the pointwise ergodic theorem is to show that  $L_2$  convergence implies  $L_1$  convergence, under some additional assumptions. Section 5.4 provides a proof that, in turn,  $L_1$  convergence implies pointwise convergence, which will conclude the proof of the pointwise ergodic theorem (in  $\text{ACA}_0$ ). For the sake of generality, though, we are going to consider general  $L_p$  spaces. Results in this section are largely adapted from [27].

The transformation  $T$  is going to be an isometry throughout. In the next claim also assume that it is multiplicative. The underlying system is  $\text{WWKL}_0$ , because the fact that  $T|f| = |Tf|$ , which is used both in Lemma 5.2.1 and Lemma 5.2.2, requires  $(\text{WWKL})$ .

**Lemma 5.2.1** ( $\text{WWKL}_0$ ) *Let  $T : L_1(X) \rightarrow L_1(X)$  be a multiplicative isometry. Then  $T$ , when restricted to  $L_p(X)$  for any  $p > 1$ , is an isometry on that space. Moreover,  $T$  takes  $L_p(X)$  to  $L_p(X)$ .*

*Proof.* Let  $f = \langle f_n \rangle$  be in  $L_1(X)$ . Observe that, since  $f_n \in C(X)$ , all powers of  $f_n$  are in  $L_p(X)$  for all  $p$ .

$$\begin{aligned} \|f\|_p^p &= \lim_n \|f_n\|_p^p = \lim_n \mu(|f_n|^p) \\ &= \lim_n \mu(|T(|f_n|^p)|) = \lim_n \mu(|(T|f_n|)^p|) \\ &= \lim_n \mu(|T(|f_n|)|^p) = \lim_n \|Tf_n\|_p^p = \|Tf\|_p^p. \end{aligned}$$

The limit in the last step exists since  $T$  is an isometry. The step before last is the consequence of  $T|f| = |Tf|$ .

For the second part of the claim, let  $f = \langle f_n \rangle \in L_p(X)$ :  $\|f_m - f_n\|_p \leq 2^{-m}$  for all  $m, n$  such that  $n > m$ . But since  $T$  is an isometry on  $L_p(X)$ ,

$$\|Tf_m - Tf_n\|_p = \|T(f_m - f_n)\|_p = \|f_m - f_n\|_p \leq 2^{-m}.$$

Therefore,  $\langle Tf_n \rangle$  is strong Cauchy with respect to the  $L_p$  norm, so  $Tf \in L_p(X)$ .  $\square$

**Lemma 5.2.2** ( $\text{WWKL}_0$ ) *If  $T$  is an isometry on  $L_p(X)$  for some  $p > 1$ , then it is also an isometry on  $L_1(X)$ .*

*Proof.* It will suffice to show that, if  $f \in C(X)$ , then  $\|Tf\| = \|f\|$ . By assumption, for every  $g \in C(X)$

$$\mu(|Tg|^p) = \mu(|g|^p).$$

Let  $g$  be such that  $|f| = g^p$ , i.e.  $g = |f|^{1/p}$ . It is easily shown that  $g$  is in  $C(X)$ . Therefore:

$$\begin{aligned} \mu(|f|) &= \mu(|g|^p) = \mu(|Tg|^p) = \mu(|T(g^p)|) \\ &= \mu(|T|f||) = \mu(|Tf|). \end{aligned}$$

□

Put together, the previous two lemmas yield the following:

**Corollary 5.2.3** (WWKL<sub>0</sub>) *If  $T$  is a multiplicative isometry on some  $L_p$ , with  $p \geq 1$ , then it is an isometry for all  $L_q$ ,  $q \geq 1$ .*

In particular,  $T$  is an isometry on  $L_1(X)$  if and only if it is an isometry on  $L_2(X)$ .

**Lemma 5.2.4** (RCA<sub>0</sub>) *Let  $T$  be a multiplicative isometry on  $L_p(X)$ . If  $\langle S_n f \rangle$  converges in  $L_q$  norm for all  $f \in L_q(X)$ , for some  $q > p$ , then  $\langle S_n f \rangle$  converges in  $L_p$  norm for all  $f \in L_p(X)$ .*

*Note:* In WWKL<sub>0</sub> there would be no need to specify where  $T$  is an isometry. This is due to the previous corollary: it is an isometry simultaneously on all  $L_p(X)$  spaces.

*Proof.* Let  $f \in L_p(X)$  be represented by  $\langle f_k \rangle$ . Then, for each  $k$ ,  $f_k \in S(X) \subseteq L_q(X)$ , and therefore for each  $k$ ,  $\langle S_n f_k \mid n \in \mathbb{N} \rangle$  converges in  $L_q(X)$  to some  $g_k \in L_q \subseteq L_p$ .

Find an estimate for  $\|g_k - g_m\|_p$  for any  $k, m$ , where  $k < m$ . The goal is to show that the sequence  $\langle g_k \rangle$  is strong Cauchy in  $L_p(X)$  and thus convergent to an element in the space. This will precisely be the limit of  $\langle S_n f \rangle$ .

For  $k < m$ ,  $\|S_n f_k - S_n f_m\|_p = \|S_n(f_k - f_m)\|_p \leq \|f_k - f_m\|_p \leq 2^{-k}$ .

Choose  $n$  so that  $\|S_n f_k - g_k\|_q < 2^{-k}$ ,  $\|S_n f_m - g_m\|_q < 2^{-k}$  (this can be done recursively, since the limit exists). Because  $p < q$ , the above inequalities also hold in  $L_p$  norm, based on the comment after Lemma 4.3.3 on page 59. Then

$$\begin{aligned} \|g_k - g_m\|_p &\leq \|g_k - S_n f_k\|_p + \|S_n f_k - S_n f_m\|_p + \|S_n f_m - g_m\|_p \\ &\leq 2^{-k} + 2^{-k} + 2^{-k} = 3 \cdot 2^{-k}. \end{aligned}$$

This is clearly a (strong) Cauchy sequence and since the space is complete, the proof is done.  $\square$

**Corollary 5.2.5** (RCA<sub>0</sub>) *Let  $T$  be an isometry on  $L_1(X)$ . Then  $L_2$  convergence of  $\langle S_n f \rangle$  for all  $f \in L_2(X)$  implies  $L_1$  convergence of  $\langle S_n f \rangle$  for all  $f \in L_1(X)$ .*

By Theorem 3.1.3, arithmetic comprehension proves  $L_2$  convergence of  $\langle S_n \rangle$ . The following holds:

**Corollary 5.2.6** (ACA<sub>0</sub>) *For every  $f \in L_1(X)$ ,  $\langle S_n f \rangle$  converges in  $L_1$  norm.*

The converse to Lemma 5.2.4 holds. It is not relevant to the proof of the ergodic theorem, however.

**Lemma 5.2.7** (RCA<sub>0</sub>) *If  $T$  is a multiplicative isometry on  $L_q(X)$ , and if  $\langle S_n f \rangle$  converges in  $L_p$  norm for all  $f \in L_p(X)$ , then  $\langle S_n f \rangle$  converges in  $L_q$  norm for all  $f \in L_q(X)$  whenever  $q > p$ .*

*Proof.* We will show that  $\langle S_n f \rangle$  is a Cauchy sequence in  $L_q(X)$ .

$$\|S_n f - S_m f\|_q \leq \|S_n f - S_n f_k\|_q + \|S_n f_k - S_m f_k\|_q + \|S_m f_k - S_m f\|_q,$$

and the first and third term are dominated by  $2^{-k}$ . As for the middle term, note that  $f_k \in S(X)$ , hence  $|f_k| \leq M_k$  for some real number  $M_k$ . Then  $|T^i f_k| \leq M_k$  for all  $i$ , which implies that for all  $n$ , also  $|S_n f_k| \leq M_k$  and

$$\begin{aligned} \|S_n f_k - S_m f_k\|_q^q &= \mu(|S_n f_k - S_m f_k|^q) \\ &= \mu(|S_n f_k - S_m f_k|^p |S_n f_k - S_m f_k|^{q-p}) \\ &\leq (2M_k)^{q-p} \|S_n f_k - S_m f_k\|_p^p, \end{aligned}$$

or

$$\|S_n f_k - S_m f_k\|_q \leq (2M_k)^{1-p/q} \|S_n f_k - S_m f_k\|_p^{p/q}.$$

The fact that  $\langle S_n f_k \rangle$  converges in  $L_p(X)$  implies that  $\langle S_n f_k \rangle$  is Cauchy in  $L_p(X)$ , and therefore for every  $l$ , and  $m$  and  $n$  large enough,  $\|S_n f_k - S_m f_k\|_p \leq 2^{-l}$ , therefore

$$\|S_n f_k - S_m f_k\|_p \leq (2M_k)^{1-p/q} 2^{-lp/q},$$

and it is not difficult to see that the sequence  $\langle S_n f_k \rangle$  is strong Cauchy in  $L_q(X)$ , as is  $\langle S_n f \rangle$ , and is therefore convergent.  $\square$

The next step is passing from convergence in norm to pointwise convergence. This requires the maximal ergodic theorem.

### 5.3 The Maximal Ergodic Theorem

The standard proof of the maximal ergodic theorem presents a problem similar to that encountered when trying to formalize the standard proof of the pointwise ergodic theorem. In this case the problem is resolved by adapting a constructive proof by Garsia [15], which (unlike the standard one) makes no use of inverse images of sets.

Recall that if a transformation is multiplicative, it is also nonnegative (in  $WWKL_0$  and stronger systems).

**Theorem 5.3.1 (ACA<sub>0</sub>)** *Let  $T$  be a multiplicative isometry on  $L_1(X)$ ,  $f \in L_1(X)$ , and  $M = \{x \mid \sup_n S_n f(x) > 0\}$ . Then  $\mu(f\chi_M) \geq 0$ .*

*Proof.* Proposition 4.5.4 shows that  $\sup_n S_n f$  is an integrable function, and by Lemma 4.4.2,  $M$  is an integrable set.

Define

$$A_0 f = 0, \quad A_n f = \sum_{k=0}^{n-1} T^k f,$$

and notice that  $S_n f = \frac{1}{n} A_n f$  for  $n \geq 1$ . Also note that

$$\{x \mid \sup_n S_n f(x) > 0\} = \{x \mid \sup_n A_n f(x) > 0\},$$

so consider the latter set.

Fix  $N \in \mathbb{N}$ . Let  $h_N = \max_{0 \leq n \leq N} A_n f$ . Then  $h_N \in L_1(X)$  and  $h_N \geq 0$ . Let  $M_N = \{x \mid h_N(x) > 0\}$ . By Lemma 4.4.2,  $M_N$  is an integrable set and  $M = \bigcup_N M_N$ . Since  $M_N$  is integrable,  $\chi_{M_N} \in L_1(X)$ .

For all  $n$ ,  $0 \leq n \leq N$ ,

$$\begin{aligned} h_N \geq A_n f &\rightarrow Th_N \geq TA_n f \\ &\rightarrow Th_N + f \geq A_{n+1} f, \end{aligned}$$

therefore,

$$\begin{aligned} Th_N + f &\geq \max_{0 \leq n \leq N} A_{n+1} f \\ &\geq \max_{1 \leq n \leq N} A_n f. \end{aligned}$$

This implies that  $Th_N \chi_{M_N} + f \chi_{M_N} \geq \max_{1 \leq n \leq N} A_n f \chi_{M_N}$ , and this fact is true in  $RCA_0$ , assuming  $\chi_{M_N}$  is an integrable function. The next step, however, requires  $WWKL_0$ :

$$\max_{1 \leq n \leq N} A_n f \chi_{M_N} = \max_{0 \leq n \leq N} A_n f \chi_{M_N} = h_N \chi_{M_N},$$

because  $h_N(x) > 0$  on  $M_N$ , so the maximum of  $A_n f$  is not attained at  $n = 0$ . Therefore,

$$f\chi_{M_N} \geq h_N\chi_{M_N} - Th_N\chi_{M_N}.$$

After integrating this becomes

$$\mu(f\chi_{M_N}) \geq \mu(h_N\chi_{M_N}) - \mu(Th_N\chi_{M_N}) \quad (5.1)$$

$$= \mu(h_N) - \mu(Th_N\chi_{M_N}) \quad (5.2)$$

$$\geq \mu(h_N) - \mu(Th_N). \quad (5.3)$$

The equality in (5.2) follows from the fact that  $h_N = h_N\chi_{M_N} + h_N\chi_{\bar{M}_N}$  and  $h_N = 0$  on  $X \setminus M_N$ , therefore  $h_N = h_N\chi_{M_N}$  while (5.3) is a consequence of nonnegativity of  $T$ :  $Th_N \geq 0$  implying that  $Th_N\chi_{M_N} \leq Th_N$  (the former is provable in WWKL<sub>0</sub>, the latter in RCA<sub>0</sub>).

As  $T$  is norm preserving, and  $h_N$  and  $Th_N$  are nonnegative,  $\mu(h_N) = \|h_N\| = \|Th_N\| = \mu(Th_N)$ , and  $\mu(f\chi_{M_N}) \geq 0$ .

Now pass to the limit. It is easy to show that  $\chi_{M_N} \rightarrow \chi_M$  and  $f\chi_{M_N} \rightarrow f\chi_M$  in  $L_1(X)$ , which implies that  $\mu(f\chi_{M_N}) \rightarrow \mu(f\chi_M)$ . Consequently,  $\mu(f\chi_M) \geq 0$ .  $\square$

If  $G = \{x \mid \sup_n S_n f(x) > \lambda\}$ , then it immediately follows from the proof above that  $\mu(\chi_G) \leq \frac{1}{\lambda}\mu(f\chi_G)$ .

Although this proof is constructive from the standpoint of Bishop-style mathematics, this is not the case in second-order arithmetic. First of all, the claims that  $\sup_n S_n f$  is an integrable function, and  $M$  and  $M_N$  for all  $N$  are integrable sets require arithmetic comprehension. The operation of taking the limit over  $N$  can be performed in RCA<sub>0</sub> with the knowledge that  $\chi_M$  and  $\chi_{M_N}$  are integrable functions. Even if we suppose these facts beforehand, the proof of the maximal ergodic theorem still does not formalize in RCA<sub>0</sub> because pointwise properties of the function  $h_N$  are used in the proof.

We will need the following modification of the maximal ergodic theorem for the third proof of the pointwise ergodic theorem: we need to show that the conclusion of the theorem still holds if we restrict ourselves to an invariant subset of the space ( $A$  is invariant if  $T\chi_A = \chi_A$ ).

**Corollary 5.3.2 (ACA<sub>0</sub>)** *Let  $T$  be a multiplicative isometry on  $L_1(X)$ ,  $f \in L_1(X)$ , and  $A$  an invariant set. If  $M = \{x \mid (\sup_n S_n f(x))\chi_A > 0\}$ , then  $\mu(f\chi_A\chi_M) \geq 0$ .*

*Proof.* Apply 5.3.1 to the integrable function  $f\chi_A$ . It suffices to show that  $\{x \mid (\sup_n S_n f)\chi_A > 0\} = \{x \mid \sup_n S_n(f\chi_A) > 0\}$ , for then the statement immediately follows. This is where the invariance of  $A$  is used: since

$T^k(f\chi_A) = T^k f \cdot T^k \chi_A = T^k f \cdot \chi_A$  for all  $k$ ,  $S_n(f\chi_A) = (S_n f)\chi_A$  and the two sets are equal.  $\square$

As above, if  $G = \{x \mid \sup_n S_n f(x) > \lambda\}$ , then  $\mu(\chi_{G \cap A}) \leq \frac{1}{\lambda} \mu(f\chi_{G \cap A})$ .

## 5.4 First Proof of the Main Theorem

Before proving the main theorem, one last auxiliary statement is needed.

**Lemma 5.4.1** (ACA<sub>0</sub>) *If  $\langle f_k \rangle$  is a sequence of functions in  $L_1(X)$  that converges in norm to  $f \in L_1(X)$ , and if for all  $k$ ,  $\lim_n S_n f_k$  exists a.e, then  $\langle S_n f \rangle$  converges a.e.*

*Note:* The above statement can be interpreted as saying that the set of all functions for which  $S_n f$  converges a.e. is  $L_1$  closed.

*Proof.* Let  $G$  be the null set on which neither  $f$ ,  $S_n f$  for any  $n$  nor  $\lim_n S_n f_k$  for any  $k$  are defined (a countable union of null  $G_\delta$  sets is a null  $G_{\delta\sigma}$  set, so integrable). Due to regularity of measure, there is no loss in assuming that  $G$  is a null  $G_\delta$  set. If  $x$  is outside of this set, and  $n, m, k \in \mathbb{N}$  are arbitrary, then

$$\begin{aligned} |S_n f(x) - S_m f(x)| &\leq |S_n f(x) - S_n f_k(x)| + |S_n f_k(x) - S_m f_k(x)| \\ &\quad + |S_m f_k(x) - S_m f(x)|. \end{aligned}$$

Since  $\lim_n S_n f_k(x)$  exists, there is  $N_k \in \mathbb{N}$  such that  $|S_n f_k(x) - S_m f_k(x)| < 2^{-k}$  whenever  $m, n \geq N_k$ . It remains to approximate the other two terms. Because  $|S_i f(x) - S_i f_k(x)| \leq \sup_j |S_j(f(x) - f_k(x))|$  for all  $i$ ,

$$|S_n f(x) - S_m f(x)| \leq 2 \sup_j |S_j(f(x) - f_k(x))| + 2^{-k}$$

for  $n, m \geq N_k$ . According to the maximal ergodic theorem, for every  $\lambda > 0$ ,

$$\begin{aligned} \mu(\{x \mid \sup_n |S_n(f(x) - f_k(x))| > \lambda\}) &\leq \\ \mu(\{x \mid \sup_n S_n |f(x) - f_k(x)| > \lambda\}) &\leq \frac{1}{\lambda} \|f - f_k\| \leq \frac{1}{\lambda} 2^{-k}. \end{aligned}$$

Let  $U_{lk} = \{x \mid \sup_n |S_n(f(x) - f_k(x))| > 2^{-l}\} \cup G$ . By Lemma 4.4.2,  $U_{lk}$  is integrable. Its measure does not exceed  $2^{l-k}$ , while  $|S_n f(x) - S_m f(x)| \leq 2 \cdot 2^{-l} + 2^{-k}$  for  $n, m > N_k$  and outside of  $U_{lk}$ .

Let  $\varepsilon$  be arbitrary. Choose  $l$  such that  $2^{-l+1} < \varepsilon/2$  and consider  $U = \bigcap_{k \geq l} U_{lk}$ :  $\mu(U) \leq 2^{l-k}$  for all  $k$ , thus  $\mu(U) = 0$ . If  $x \notin U$ , then  $x \notin U_{lk}$  for some  $k$ . For this value  $\sup_n |S_n(f(x) - f_k(x))| \leq 2^{-l}$  and

$$|S_n f(x) - S_m f(x)| \leq 2 \cdot 2^{-l} + 2^{-k} < \varepsilon$$

for  $m, n \geq N_k$ .

Now for each  $i$  let  $U_i$  be the set  $U$  which corresponds to  $\varepsilon = 2^{-i}$  above. Define  $V = \bigcup_i U_i$ . As a union of null sets,  $V$  is itself a null set, at a low level in the Borel hierarchy. We can therefore appeal to regularity of measure once again, and as  $V$  is contained in a null  $G_\delta$  set, it may be assumed that  $V$  itself is a  $G_\delta$  set. For  $x \notin V$ ,  $\langle S_n f(x) \rangle$  is a Cauchy sequence, thus convergent, which by definition means that  $\langle S_n f \rangle$  converges a.e.  $\square$

**Theorem 5.4.2 (ACA<sub>0</sub>) (Pointwise ergodic theorem)** *For every multiplicative isometry  $T$  and for every  $f \in L_1(X)$ ,  $\langle S_n f \rangle$  converges a.e. to an integrable function  $\hat{f}$ . In addition,  $\mu(\hat{f}) = \mu(f)$  and  $\hat{f}$  is invariant.*

*Proof.* Given  $f \in L_1(X)$ ,  $f = \langle f_n \rangle$ , where  $f_n \in S(X)$  for all  $n$ . Since  $f_n \in L_2(X)$ , by Theorem 3.1.2, ACA<sub>0</sub> proves that  $f_n = f_n^M + f_n^N$ , (follows from the mean ergodic theorem) where  $f_n^M$  and  $f_n^N$  are in  $L_2(X)$  and therefore integrable. Then  $T f_n^M = f_n^M$  and  $f_n^N$  is represented by the sequence  $\langle g_{nk} - T g_{nk} \rangle$ , with  $g_{nk} \in S(X)$  for all  $k$ .  $S_m f_n^M$  converges pointwise to  $f_n^M$ , as  $S_m f_n^M(x) = f_n^M(x)$  for all  $x$  for which it is defined.

For any function of the form  $g - Tg$ , if  $g \in C(X)$  then  $|g(x)| \leq K$  for some  $K$ , and because  $T$  preserves  $L_{1,\infty}$  norm

$$|S_n g(x)| = \frac{1}{n} |g(x) - T^n g(x)| \leq 2K/n \rightarrow 0.$$

Thus,  $S_n g$  converges pointwise to 0. As  $f_n^N$  is the  $L_1$  limit of functions of the above form (an  $L_2$  limit is also an  $L_1$  limit), according to Lemma 5.4.1,  $S_m f_n^N$  is also pointwise convergent, and, as a consequence,  $S_m f_n$  is pointwise convergent. But, once again, by the same lemma, as  $f$  is the  $L_1$  limit of the sequence  $f_n$ ,  $S_n f$  converges a.e. At the same time, it was proved above that  $\langle S_n f \rangle$  converges in  $L_1$  to an integrable function  $\hat{f}$ , which by Lemma 4.4.4 implies that  $\lim_n S_n f(x) = \hat{f}(x)$  a.e. This establishes the first part of the claim. That  $\mu(\hat{f}) = \mu(f)$  follows directly from  $\mu(\hat{f} - f) = \mu(\hat{f} - S_n f) + \mu(S_n f - f)$ , and  $\mu(S_n f - f) = 0$  while, since  $S_n f \rightarrow \hat{f}$  in norm,  $\|\hat{f} - S_n f\| \rightarrow 0$  thus  $\mu(\hat{f} - S_n f) \rightarrow 0$ .

To prove invariance of  $\hat{f}$ , it suffices to show that  $\|T\hat{f} - \hat{f}\| = 0$ . For all  $n \in \mathbb{N}$ ,

$$\|T\hat{f} - \hat{f}\| \leq \|T\hat{f} - TS_n f\| + \|TS_n f - S_n f\| + \|S_n f - \hat{f}\|.$$

Consider each of the terms on the right-hand side of the inequality. Fix  $\varepsilon > 0$ . Because  $\|\hat{f} - S_n f\| \rightarrow 0$ , there exists  $N_1(\varepsilon)$  such that  $\|T\hat{f} - TS_n f\| = \|\hat{f} - S_n f\| < \varepsilon$  for  $n > N_1(\varepsilon)$ . Similarly, there exists  $N_2(\varepsilon)$  such that  $\|TS_n f - S_n f\| = \frac{1}{n}\|T^n f - f\| \leq \frac{2}{n}\|f\| < \varepsilon$  for  $n > N_2(\varepsilon)$ . For  $n > \max\{N_1, N_2\}$ ,  $\|T\hat{f} - \hat{f}\| < 3\varepsilon$ . Since  $\varepsilon$  is arbitrary, this concludes the proof.  $\square$

We showed that  $L_1$  convergence of the sequence  $\langle S_n f \rangle$  implies its pointwise convergence. In fact, the reverse also holds. This follows directly from Lemma 4.4.5. Since  $S_n f \rightarrow \hat{f}$  a.e. and  $\|S_n f\| = \|\hat{f}\|$ , we have  $S_n f \rightarrow \hat{f}$  in  $L_1(X)$ .

## 5.5 Second Proof

We now proceed to the second proof of the pointwise ergodic theorem. The advantage this proof has over the first one is that it requires fewer assumptions on  $T$ . It is only required that it be a measure preserving (and therefore nonnegative) linear operator. The resulting proof, however, is longer and somewhat less intuitive. It also relies heavily on arithmetic comprehension. The goal is to prove that  $\liminf_n S_n f$  and  $\limsup_n S_n f$  coincide. To even begin talking about these notions (suprema, infima, limits), arithmetic comprehension is necessary. Most proofs in this section are formalized in  $\text{ACA}_0$ .

For the sake of clarity, the proof is divided into a number of lemmas. We retain the notation from the previous sections.

Since any function  $f$  can be written as  $f = f^+ - f^-$ , and  $T$  is linear, it can safely be assumed that  $f \geq 0$ . Consider the full  $F_\sigma$  set  $X$  on which  $f$ , along with  $T^k f$  for all  $k$ , is defined and nonnegative.

Recall that for each  $n$ ,  $\|S_n f\| \leq \|f\|$ . Set  $\bar{f}(x) = \limsup_n S_n f(x)$  and  $\underline{f}(x) = \liminf_n S_n f(x)$  for  $x \in X$ . By virtue of arithmetic comprehension, both of these pointwise limits exist. Furthermore, by Proposition 4.5.4,  $\bar{f}$  and  $\underline{f}$  are integrable.

Fix  $\varepsilon > 0$ . Let

$$\psi(x, n) \equiv (n = \min\{k \geq 1 \mid S_k f(x) \geq \bar{f}(x) - \varepsilon\}).$$



Then  $\psi(x, n)$  is an arithmetic formula which holds if the first average that comes within  $\varepsilon$  of  $\bar{f}(x)$  is  $S_n f(x)$ .

Next, define a formula  $\varphi$  so that

$$\begin{aligned} \varphi(x, m, n) \quad \leftrightarrow \quad & \exists \langle k_1, \dots, k_m \rangle (\psi(x, k_1) \wedge \psi(T^{k_1} x, k_2) \wedge \dots \\ & \wedge \psi(T^{k_1 + \dots + k_{m-1}} x, k_m) \wedge n = k_1 + \dots + k_m). \end{aligned}$$

In other words, once an average is reached that is within  $\varepsilon$  from  $\bar{f}$ , we start “counting” over, until another such average is reached. If  $\varphi(x, m, n)$  holds, then by term with index  $n$  this has happened  $m$  times.

We will write  $n = R(x)$  when  $\psi(x, n)$  holds and  $n = R_m(x)$  when  $\varphi(x, m, n)$  holds. Both  $R(x)$  and  $R_m(x)$  exist a.e. and  $S_{R_m(x)} f(x) \geq \bar{f}(x) - \varepsilon$  for all  $m$ .

First consider a restricted case. In the following lemma use the assumption that  $T$  is nonnegative. So far in this section, arithmetic comprehension has been necessary for all considerations, but if we assume that  $R_m(x)$  is defined a.e. for all  $m$  and that  $\bar{f} \in L_1(X)$ , this lemma can be proved in WWKL<sub>0</sub>.

**Lemma 5.5.1** (WWKL<sub>0</sub>) *If  $\bar{f} \in L_1(X)$ , and there exist  $R^*$  and  $M$  in  $\mathbb{N}$  such that  $R(x) \leq R^*$  and  $\bar{f}(x) \leq M$  a.e, then  $\mu(\bar{f}) \leq \mu(f) + 2\varepsilon$ .*

*Proof.* Pick  $L$  such that  $\frac{MR^*}{L} < \varepsilon$  and fix  $x$  in the domain of  $\bar{f}$  and for which  $R_m(x)$  is defined for all  $m$ . There exists  $n$  such that  $R_n(x) \leq L < R_{n+1}(x)$  and

$$\begin{aligned} L \cdot (\bar{f}(x) - \varepsilon) & \leq R_{n+1}(x) \cdot (\bar{f}(x) - \varepsilon) \\ & = R_n(x) \cdot (\bar{f}(x) - \varepsilon) + (R_{n+1}(x) - R_n(x)) \cdot (\bar{f}(x) - \varepsilon) \\ & \leq R_n(x) S_{R_n} f(x) + MR^* \leq LS_L f(x) + MR^*. \end{aligned}$$

The inequality  $R_{n+1}(x) - R_n(x) \leq R^*$  is deduced from

$$R_{n+1}(x) - R_n(x) \leq R_{n+1}(x) = R(R_n(x)) \leq R^*.$$

In the last step nonnegativity of  $T$  was used to conclude that  $S_{R_n} f \leq S_L f$ .

Now divide by  $L$ . Since  $\bar{f}(x) - \varepsilon \leq S_L f(x) + \frac{MR^*}{L}$  a.e, by Proposition 4.4.1:

$$\mu(\bar{f} - \varepsilon) \leq \mu(f_L) + \frac{MR^*}{L} \leq \frac{1}{L} \sum_{k=0}^{L-1} \mu(T^k f) + \varepsilon = \mu(f) + \varepsilon.$$

Use the fact that  $T$  is measure preserving in the last step.

Finally, because  $\mu(\bar{f} - \varepsilon) = \mu(\bar{f}) - \mu(\varepsilon) = \mu(\bar{f}) - \varepsilon$  (since  $\mu(1) = 1$ ) it follows that  $\mu(\bar{f}) \leq \mu(f) + 2\varepsilon$ .  $\square$

Before proceeding to the general case, we prove the following statement.

**Lemma 5.5.2** (ACA<sub>0</sub>) *The set  $W = \{x \mid R(x) \geq K\}$  is integrable for all  $K \in \mathbb{N}$ .*

*Proof.* This fact doesn't follow immediately from Lemma 4.4.2 because  $R(x)$  is not defined as a function.

The formula  $R(x) \geq K$  is an abbreviation for  $\psi(x, n) \wedge n \geq K$ . Let  $W_n = \{x \mid S_n(x) - \bar{f}(x) + \varepsilon \geq 0\}$  and  $Y_n = W_n \setminus \bigcup_{i=1}^{n-1} W_i$ . Then  $x$  is in  $W_n$  if and only if  $S_n(x)$  is within  $\varepsilon$  from  $\bar{f}(x)$  and is in  $Y_n$  if and only if it is there for the first time:  $x \in Y_n \leftrightarrow R(x) = n$ . Hence,

$$W = \bigcup_{n=K}^{\infty} Y_n = \bigcup_{n=K}^{\infty} W_n \setminus \bigcup_{n=1}^{K-1} W_n.$$

Each  $W_i$  is integrable by Lemma 4.4.2, as is  $W$ , because infinite unions of integrable sets are integrable, as are differences of integrable sets.  $\square$

To conclude the proof, consider an arbitrary  $f$ , and “cut off”  $\bar{f}$  at some  $M$ , i.e. make it essentially bounded to reduce to the previous case. Taking limits will give us the desired result.

**Lemma 5.5.3** (ACA<sub>0</sub>) *For any  $f \in L_1(X)$ ,  $\mu(\bar{f}) \leq \mu(f)$ .*

*Proof.* Fix  $M \in \mathbb{N}$  and set  $\bar{f}_M = \min(\bar{f}, M)$ . Choose  $R^* \in \mathbb{N}$  such that  $\mu(\{x \mid R(x) \geq R^*\}) \leq \frac{\varepsilon}{M}$ . This is possible because  $R(x)$  is finite a.e. and  $\mu(X) = 1$ .

Let  $\psi_M(x, n) \equiv (n = \min\{k \geq 1 \mid S_n f(x) \geq \bar{f}_M - \varepsilon\})$ . As before, when  $\psi_M(x, n)$  holds, denote  $n = R_M(x)$ . Note that  $R_M(x) \leq R(x)$  since  $\bar{f}_M \leq \bar{f}$ . Define  $\tilde{f}$  as

$$\tilde{f}(x) = \begin{cases} f(x) & R(x) \leq R^* \\ \max(f(x), M) & R(x) > R^*, \end{cases}$$

and define  $\tilde{R}$  with

$$\tilde{R}(x) = \begin{cases} R_M(x) & R(x) \leq R^* \\ 1 & R(x) > R^*. \end{cases}$$

Then  $\tilde{f} \in L_1(X)$  because  $\tilde{f} = f\chi_{R \leq R^*} + \max(f, M)\chi_{R > R^*}$  a.e. and both  $\{R \leq R^*\}$  and  $\{R > R^*\}$  are integrable sets and  $\max(f, M)$  is an integrable

function. Integrability of  $\tilde{f}$  follows from the comment in Section 4.4 on page 60. In addition,  $\tilde{f} \geq 0$ .

The following inequality holds:

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

a.e.

There are two cases. If  $R(x) \leq R^*$ , then  $\tilde{R}(x) = R_M(x)$  and  $\tilde{f}(x) = f(x)$  and

$$\frac{1}{R_M} \sum_{k=0}^{R_M-1} T^k f \geq \bar{f}_M - \varepsilon$$

by definition.

The second case occurs when  $R(x) > R^*$ . Then  $\tilde{R} = 1$  and

$$\frac{1}{1} \sum_{k=0}^0 T^k \tilde{f} = \tilde{f} = \max(f, M) \geq M \geq \bar{f}_M - \varepsilon.$$

The conditions of Lemma 5.5.1 are now satisfied. This lemma applies to  $\bar{f}_M$  in place of  $\bar{f}$  and  $\tilde{f}$  in place of  $f$  and thus

$$\mu(\bar{f}_M) \leq \mu(\tilde{f}) + 2\varepsilon.$$

At the same time,

$$\begin{aligned} \mu(\tilde{f}) &= \mu(f\chi_{R \leq R^*}) + \mu(\max(f, M)\chi_{R > R^*}) \\ &\leq \mu(f\chi_{R \leq R^*}) + \mu((M + f)\chi_{R > R^*}) \\ &\leq \mu(f) + M \frac{\varepsilon}{M} = \mu(f) + \varepsilon, \end{aligned}$$

thus  $\mu(\bar{f}_M) \leq \mu(f) + 3\varepsilon$ .

Since  $\varepsilon$  was arbitrary,  $\mu(\bar{f}_M) \leq \mu(f)$ . Since  $\bar{f} = \lim_M \bar{f}_M$ ,

$$\mu(\bar{f}) = \mu(\lim_M \bar{f}_M) = \lim_M \mu(\bar{f}_M) \leq \mu(f).$$

The inequality is the consequence of the proof above. The limit commutes with the integral due to the weaker version of monotone convergence theorem, provable in WWKL<sub>0</sub>, as  $\bar{f}_M$  is an increasing sequence of integrable functions that converges to an integrable function  $\bar{f}$ .  $\square$

**Theorem 5.5.4 (ACA<sub>0</sub>) (Pointwise ergodic theorem)** *Let  $T : L_1(X) \rightarrow L_1(X)$  be a nonnegative measure preserving transformation and  $f \in L_1(X)$ . Then the sequence of averages  $S_n f = \sum_{k=0}^{n-1} T^k f$  converges for almost all  $x \in X$ .*

*Proof.* The following chain of inequalities holds:

$$\mu(f) \leq \mu(\underline{f}) \leq \mu(\overline{f}) \leq \mu(f).$$

The middle inequality is trivial, the third is Lemma 5.5.3, while the first is analogous to the first.

To see that this proves the theorem, note that

$$\mu(\overline{f} - \underline{f}) \leq \mu(f - f) = 0.$$

By Proposition 4.4.1,  $\|f\| = 0$  iff  $f = 0$  a.e. Since  $f$  is nonnegative,  $\mu(\overline{f} - \underline{f}) = 0$  implies  $\overline{f} = \underline{f}$  outside of a null  $G_\delta$  set  $G$ . For all  $x$  outside of  $G$ , due to nested interval convergence,  $\overline{f}(x) = \underline{f}(x) = \lim_n S_n f(x)$ . This implies that the sequence  $\langle S_n f \rangle$  converges to an integrable function  $\hat{f}$  a.e. as required.

## 5.6 Third Proof

In this section we consider the standard proof of the pointwise ergodic theorem. In the classical mathematical setting this proof is short and straightforward. In our framework, however, it becomes complicated, providing an example of the difficulties in dealing with subsets of a measure space in the context of second-order arithmetic. For that reason, this proof is presented last. To simplify matters, the underlying space  $X$  is taken to be  $[0, 1]$ , though it is likely that the argument easily generalizes to an arbitrary compact metric space equipped with a Borel measure. The main idea of the proof is to show that given  $a, b \in \mathbb{Q}$  such that  $a < b$ , the set

$$A_{a,b} = \{x \mid \liminf_n S_n f(x) < a < b < \limsup_n S_n f(x)\}$$

is invariant ( $T^{-1}(A) = A$ ), and then with the help of the maximal ergodic theorem to show that  $\mu(A_{a,b}) = 0$ . Since  $\bigcup_{a,b} A_{a,b}$  is the set of all  $x$  for which the limit of  $S_n f(x)$  does not exist, and since  $\mu(\bigcup_{a,b} A_{a,b}) = 0$ , this suffices to prove the claim. The main difficulty for us is in showing that  $A_{a,b}$  is an invariant set. In our setting this means that  $T\chi_{A_{a,b}} = \chi_{A_{a,b}}$ .

Throughout the section we are going to work under the assumption that  $T$  is a multiplicative isometry on  $L_1([0, 1])$ . Recall that a multiplicative transformation also preserves  $L_{1,\infty}$  norm.

Notice that  $A_{a,b} = B \cap C$ , where

$$B = \{x \mid \liminf_n S_n f(x) < a\},$$

and

$$C = \{x \mid b < \limsup_n S_n f(x)\},$$

and that

$$T\chi_{B \cap C} = T(\chi_B \cdot \chi_C) = T\chi_B \cdot T\chi_C.$$

If we show that  $B$  and  $C$  are both invariant,  $A_{a,b}$  will be invariant as well. As  $B$  and  $C$  have symmetric definitions, it is enough to prove the claim for only one of the two, so from now on consider only the set  $B$ .

Since  $x \in B \leftrightarrow \exists \varepsilon \forall k \exists n \geq k (S_n f(x) < a - \varepsilon)$ ,

$$B = \bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} \{x \mid S_n f(x) < a - \varepsilon\}.$$

For brevity, let

$$B_{n\varepsilon} = \{x \mid S_n f(x) < a - \varepsilon\},$$

where  $\varepsilon$  ranges over rational numbers, e.g. those of the form  $2^{-m}$  for some  $m$ .

The next thing we need to show is that  $T$ , in a sense, commutes with infinite unions and intersections, i.e. that

$$T(\chi_{\bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} B_{n\varepsilon}}) = \chi_{\bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} D_{n\varepsilon}},$$

where  $D_{n\varepsilon}$  is such that  $T\chi_{B_{n\varepsilon}} = \chi_{D_{n\varepsilon}}$ .

This fact is true in the finite case (see also comment on page 75 before Proposition 5.1.5). It is not difficult to show that, due to continuity of  $T$ , the extension to infinite unions and intersections can be made. For example, assume  $M = \bigcup_{n=1}^{\infty} M_n$  exists (meaning  $\chi_M = \lim_n \chi_{\bigcup_{k=1}^n M_k}$ ), with  $T\chi_{M_n} = \chi_{N_n}$ . Then

$$\begin{aligned} T\chi_M &= T(\chi_{\lim_n \bigcup_{k=1}^n M_k}) = T(\lim_n \chi_{\bigcup_{k=1}^n M_k}) \\ &\stackrel{(\star)}{=} \lim_n T(\chi_{\bigcup_{k=1}^n M_k}) = \lim_n \chi_{\bigcup_{k=1}^n N_k} = \chi_{\bigcup_{n=1}^{\infty} N_n} \end{aligned}$$

The step  $(\star)$  follows from continuity of  $T$ .

With this conclusion, and since

$$T\chi_B = T\chi_{\bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} B_{n\varepsilon}} = \chi_{\bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} C_{n\varepsilon}},$$

it suffices to show that  $\cup_\varepsilon \cap_k \cup_{n \geq k} B_{n\varepsilon} = \cup_\varepsilon \cap_k \cup_{n \geq k} C_{n\varepsilon}$ . To make this claim, we need to see how  $T$  acts on  $\chi_{B_{n\varepsilon}}$ . For this purpose we consider an alternative characterization of integrable functions.

**Definition 5.6.1** ( $\text{RCA}_0$ ) *A function  $f$  is called a step function if it is of the form  $\sum_{k \leq n} c_k \chi_{I_k}$ , where  $I_1, \dots, I_n$  are disjoint open intervals.*

It can be shown in  $\text{RCA}_0$  that every step function is integrable, because characteristic functions of open intervals can be defined in  $\text{RCA}_0$  (see page 46) and linear combinations of integrable functions are integrable (comment before Lemma 4.4.4).

**Lemma 5.6.2** ( $\text{RCA}_0$ ) *Every function  $f \in L_1(X)$  is the  $L_1$  limit of step functions.*

*Proof.* It is enough to show the claim for simple functions. Let  $f$  be a simple function. Let  $g_n$  be the function defined as  $\sum_{k=0}^{n-1} f(\frac{k}{n}) \chi_{I_{kn}}$ , where  $I_{kn} = (\frac{k}{n}, \frac{k+1}{n})$ . It is clear that each  $g_n$  is a step function. It can furthermore be shown that  $\langle g_n \rangle$  is a Cauchy sequence and that it converges to  $f$ .  $\square$

**Lemma 5.6.3** ( $\text{WWKL}_0$ ) *Let  $g \in L_1(X)$ . Then*

$$T\chi_{\{x \mid g(x) > 0\}} = \chi_{\{x \mid Tg(x) > 0\}}. \quad (5.4)$$

*Proof.* We will use the previous characterization of integrable functions to prove the claim. First assume that  $g$  is a characteristic function of an open interval: call it  $\chi_I$ .

This case is simple, since for all  $x$ ,  $\chi_I(x) > 0 \leftrightarrow \chi_I(x) = 1$  (if  $\chi_I(x)$  is defined, it is either 0 or 1, according to the definition of characteristic functions of intervals) and  $\{x \mid \chi_I(x) = 1\} = I$ , so the left-hand side of (5.4) becomes  $T\chi_I$ . As for the right-hand side, since  $T\chi_I$  is the characteristic function of some set  $J$ , similarly,  $\chi_J(x) > 0 \leftrightarrow \chi_J(x) = 1$  and  $\chi_{\{x \mid \chi_J(x) = 1\}} = \chi_J$ , so the two are equal.

Next assume that  $g = \sum_{k \leq n} c_k \chi_{I_k}$ . Since the intervals  $I_k$  are disjoint,

$$\{x \mid \sum_{k \leq n} c_k \chi_{I_k} > 0\} = \cup_{k \leq n} \{x \mid c_k \chi_{I_k} > 0\}.$$

Observe that, if  $T\chi_{I_k} = \chi_{J_k}$ , the sets  $J_1, \dots, J_n$  can intersect only at a null set, as

$$\begin{aligned} \chi_{J_i \cap J_k} &= \chi_{J_i} \cdot \chi_{J_k} = T\chi_{I_i} \cdot T\chi_{I_k} \\ &= T(\chi_{I_i \cap I_k}) = T(0) = 0. \end{aligned}$$

Since two sets that differ only on a null set have the same characteristic function, and since  $T$  commutes with unions in the sense specified above, it is enough to consider only  $T\chi_{\{x \mid c_k \chi_{I_k} > 0\}}$  for a fixed  $k$ .

If  $c_k \leq 0$ , then  $\{x \mid c_k \chi_{I_k} > 0\} = \emptyset$ ,  $\chi_\emptyset = 0$  and  $T(0) = 0$ . Following the same reasoning,  $\{x \mid T(c_k \chi_{I_k} > 0)\} = \{x \mid c_k \chi_{J_k} > 0\} \emptyset$ , and once again, the two sides of (5.4) coincide. If  $c_k > 0$ , then  $\{x \mid c_k \chi_{I_k} > 0\} = \{x \mid \chi_{I_k} > 0\}$  and this reduces to the previous case.

Finally, let  $g$  be represented as a limit of step functions  $\langle g_n \rangle$ . Then

$$g(x) > 0 \leftrightarrow \forall \varepsilon \exists N \forall n \geq N g_n(x) > \varepsilon,$$

and

$$\{x \mid g(x) > 0\} = \bigcap_{\varepsilon} \bigcup_N \bigcap_n \{x \mid g_n(x) > \varepsilon\}.$$

We can once again appeal to the fact that  $T$  commutes with unions and intersections. Since  $T\chi_{\{x \mid g_n(x) > \varepsilon\}} = \chi_{\{x \mid Tg_n(x) > \varepsilon\}}$  for all  $n$ , it follows that

$$\begin{aligned} T\chi_{\{x \mid g(x) > 0\}} &= T\chi_{\bigcap_{\varepsilon} \bigcup_N \bigcap_n \{x \mid g_n(x) > \varepsilon\}} \\ &= \chi_{\bigcap_{\varepsilon} \bigcup_N \bigcap_n \{x \mid Tg_n(x) > \varepsilon\}} \\ &= \chi_{\{x \mid Tg(x) > 0\}}. \end{aligned}$$

The last step follows from the fact that  $T$  is continuous.  $\square$

Now we need to see why this is enough to prove our initial claim. The lemma is stated and proved in WWKL<sub>0</sub>, with the assumption that all limits exist.

**Lemma 5.6.4** (WWKL<sub>0</sub>) *The set  $B$  is invariant.*

*Proof.* First assume that the function  $f$  is bounded. According to the previous lemma,  $T\chi_{\{x \mid S_n f(x) < a - \varepsilon\}} = \chi_{\{x \mid TS_n f(x) < a - \varepsilon\}}$  and

$$\begin{aligned} T\chi_A &= T\chi_{\{x \mid \liminf_n S_n f(x) < a\}} \\ &= T\chi_{\bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} \{x \mid S_n f(x) < a - \varepsilon\}} \\ &= \chi_{\bigcup_{\varepsilon} \bigcap_k \bigcup_{n \geq k} \{x \mid TS_n f(x) < a - \varepsilon\}} \\ &= \chi_{\{x \mid \liminf_n TS_n f(x) < a\}}, \end{aligned}$$

and it therefore suffices to show that

$$\{x \mid \liminf_n TS_n f(x) < a\} = \{x \mid \liminf_n S_n f(x) < a\}.$$

Notice that  $TS_n f(x) = S_n f(x) + \frac{T^n f(x) - f(x)}{n}$ . Since  $f$  is bounded,  $|f| \leq M_f$ , and as  $T$  preserves  $L_{1,\infty}$  norm,  $|\frac{T^n f(x) - f(x)}{n}| \leq \frac{2M}{n} \rightarrow 0$ . It is not difficult to show that, if  $\langle a_n \rangle$  and  $\langle b_n \rangle$  are two sequences of real numbers such that  $\lim_n b_n = 0$ , then  $\liminf_n (a_n + b_n) = \liminf_n a_n$ , and therefore

$$\begin{aligned} \liminf_n TS_n f(x) &= \liminf_n \left( S_n f(x) + \frac{T^n f(x) - f(x)}{n} \right) \\ &= \liminf_n S_n f(x). \end{aligned}$$

It should be clear that this proves the claim in the case  $f$  is bounded.

The general case follows, as  $f$  is represented with a sequence  $\langle f_n \rangle$  of simple functions, which are by definition bounded. More precisely, we have proved that for every  $k$ ,

$$T\chi_{\{x \mid \liminf_n S_n f_k(x) < a\}} = \chi_{\{x \mid \liminf_n S_n f_k(x) < a\}},$$

while

$$\liminf_n S_n f(x) < a \Leftrightarrow \forall \varepsilon \exists N \forall m \geq N (\liminf_n S_n f_m(x) < a - \varepsilon),$$

from which the claim follows, once again by the useful fact that  $T$  commutes with unions and intersections.  $\square$

Finally, it remains to prove the actual pointwise ergodic theorem. The proof has been adapted from [3].

**Theorem 5.6.5 (ACA<sub>0</sub>) (Pointwise ergodic theorem)** *If  $T$  is a multiplicative isometry on  $L_1([0, 1])$ , then the sequence of averages  $S_n f(x) = \frac{1}{n} \sum_{k=0}^{n-1} T^k f(x)$  converges for a.e.  $x$  in  $[0, 1]$ .*

*Proof.* As indicated at the beginning of the section, we need to show that for all rational numbers  $a$  and  $b$  with  $a < b$ ,  $\mu(A_{a,b}) = 0$ . Fix  $a$  and  $b$ . We showed that  $A_{a,b}$  is an invariant set. Furthermore, if  $M = \{x \mid \sup_{n \geq 1} S_n f(x) > b\}$ , then  $A_{a,b} = A_{a,b} \cap M$  and by Corollary 5.3.2 of the maximal ergodic theorem,

$$b\mu(\chi_{A_{a,b}}) \leq \mu(f\chi_{A_{a,b}}).$$

Similarly,

$$\mu(f\chi_{A_{a,b}}) \leq a\mu(\chi_{A_{a,b}}),$$

and

$$b\mu(\chi_{A_{a,b}}) \leq a\mu(\chi_{A_{a,b}}),$$



which, as  $a < b$ , is only possible when  $\mu(\chi_{A_{a,b}}) = 0$ , which means that  $A_{a,b}$  is a null set and

$$\mu(\{x \mid \lim_n S_n f(x) \text{ does not exist}\}) = \mu(\cup_{a,b \in \mathbb{Q}} A_{a,b}) = 0,$$

which implies that the limit exists a.e. □

## 5.7 Reversal

The introduction contained Bishop's explanation as to why the ergodic theorem of Birkhoff is not constructive. He gave an example of two vessels between which a leak may or may not exist. Imagine now infinitely many such pairs, each half the size of the previous one, so that they all fit on a finite line, and associate to the  $n^{\text{th}}$  pair the  $n^{\text{th}}$  Turing machine is some standard ordering. A leak exists iff the corresponding Turing machine halts. Knowing the behavior of the system after a sufficient amount of time provides the solution to the halting problem, and we know that is not constructive. In fact, it is equivalent to arithmetic comprehension. Formalizing this heuristic argument will provide us with the reversal of the pointwise ergodic theorem. Recall that the Turing jump of  $Z \subseteq \mathbb{N}$  is  $\{x \mid \exists y \theta(x, y, Z)\}$ , where  $\theta$  is  $\Delta_0$  and  $\exists y \theta(x, y, Z)$  is a complete  $\Sigma_1$  formula.

The transformation  $T$  defined below will preserve measure and norm, and will be multiplicative and nonnegative (recall that in  $\text{RCA}_0$  some of these notions may differ). This will imply that the equivalence between the pointwise ergodic theorem and  $(ACA)$  will hold regardless of the definition, regardless of which proof of the theorem we consider (recall that the second proof uses slightly different assumptions from the other two). Having said this, we take the statement of the theorem to be the one below, with the understanding that even with modified assumptions, the reversal works.

If  $X$  is a compact complete separable metric space, and  $T$  a multiplicative isometry on  $L_1(X)$ , then the sequence of averages  $S_n f(x) = \frac{1}{n} \sum_{k=1}^{n-1} T^k f(x)$  converges a.e. to an invariant function  $\hat{f} \in L_1(X)$  such that  $\mu(\hat{f}) = \mu(f)$ .

**Theorem 5.7.1** ( $\text{RCA}_0$ ) *The following are equivalent:*

1. *Pointwise ergodic theorem.*
2. *(ACA).*

*Proof.* It only remains to prove the reversal. Let the space  $X$  be  $[0, 1]$  with the standard metric. It is compact as required and the Borel measure on this space coincides with the Riemann integral. Define a measure preserving transformation  $T$  on  $L_1(X)$  in the following way:

We split  $[0, 1]$  into intervals  $I_k = [x_k, x_{k+1})$ , where  $x_k = 1 - 2^{-k}$  and  $k \geq 0$ , so that  $I_k$  has length  $2^{-(k+1)}$ . Define

$$(a_k)_n = \begin{cases} 2^{-(k+m)} & m \leq n \wedge \theta(k, m, Z) \wedge \forall l < m \neg \theta(k, l, Z) \\ 0 & \text{otherwise.} \end{cases}$$

It is immediate that each  $a_k$  is a computable real number and that  $a_k$  is equal to 0 if and only if  $\neg \exists m \theta(k, m, Z)$ .

Next, define  $T$  on simple functions. If  $f \in S(X)$  and  $x \in [x_{k-1}, x_k)$ , define

$$Tf(x) = \begin{cases} f(x + a_k) & x + a_k < x_k \\ f(x + a_k - 2^{-k}) & x + a_k > x_k. \end{cases}$$

Note that  $Tf$  is not defined when  $x + a_k = x_k$ , but since there are only countably many such points, this presents no difficulty.

We will show that  $T$  satisfies all the properties previously associated with it. First show that  $T$  is well-defined, i.e. that  $Tf \in L_1([0, 1])$ . This is due to the fact that  $Tf$  is a piecewise continuous function, with at most countably many jump discontinuities (they can occur at  $x_{k-1}$ ,  $x_{k-1} + a_k$  and  $x_k$  for all  $k$ ), which is moreover bounded above; by Lemma 4.5.2,  $Tf$  is an integrable function.

Next show that  $T$  is measure preserving on  $S(X)$ . Let  $f$  be a simple function. Because  $\mu(f) = \int_0^1 f(x)dx = \sum_k \int_{x_{k-1}}^{x_k} f(x)dx$ , it is enough to show that  $\int_{x_{k-1}}^{x_k} Tf(x)dx = \int_{x_{k-1}}^{x_k} f(x)dx$ , but this fact follows immediately from invariance of the Riemann integral under translation. The same argument can be used to show that  $T$  is an isometry.

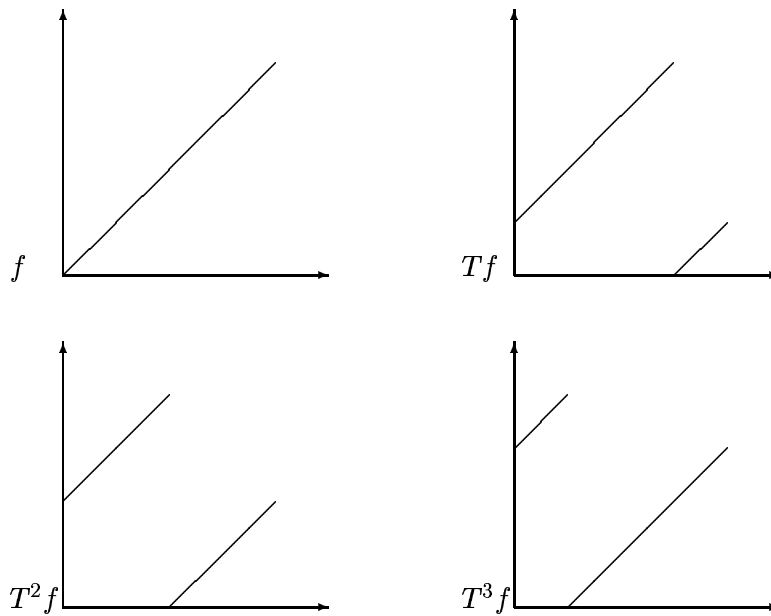
Showing that  $T$  is nonnegative is also straightforward, and it is clear that if  $f \in S(X)$ ,  $Tf \in L_{1,\infty}$ , as the maximum of  $f$  is preserved. It remains to show that if  $f, g \in S(X)$ , then  $T(fg) = Tf \cdot Tg$ . That  $T(fg)(x) = Tf(x)Tg(x)$  whenever all the quantities are defined is immediate. How about the strong product of  $Tf$  and  $Tg$ ? We know that this product exists and is unique. By a previous remark, it has to coincide with the pointwise product, which is  $T(fg)$ .

Now extend the definition to all of  $L_1([0, 1])$ . Because  $T$  is an isometry on simple functions, this will imply that  $T$  is well-defined on the entire space.

Pointwise properties of integrable functions are in general not available in  $\text{RCA}_0$ . However, the only choice for  $f$  of interest to us is that of the

identity function. In this case, we will be able to analyze both  $Tf$  and  $S_n f$ . The definition of  $T$  implies that  $Tf(x)$  remains in the same subinterval as  $x$ , as do all the averages  $S_n f(x)$ . If  $a_k = 0$ ,  $T$  does not change  $f$ , hence  $S_n f = f$  on  $[x_{k-1}, x_k)$  for all  $n$ . If  $a_k \neq 0$ , each iteration moves  $f(x)$ , and regardless of where in the interval  $x$  is,  $T^{2^m} f = f$ .

Consider the graphical representation of this. For the sake of convenience, assume  $k = 0$  and  $m = 2$ .

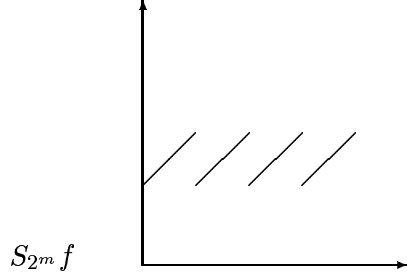


Algebraically, it can be shown that

$$\begin{aligned} S_{2^m} f(x) &= \frac{1}{2^m} [x' + (x' + a_k) + (x' + 2a_k) + \cdots + (x' + (2^m - 1)a_k)] \\ &= x' + \frac{2^m - 1}{2} a_k = x' + 2^{-(k+1)} - 2^{-(k+m+1)}. \end{aligned}$$

where  $x' = x - i2^{-m}$ , and  $i$  is the largest integer such that  $x - i2^{-m} \geq x_{k-1}$ , that is,  $x'$  is the leftmost iteration of  $x$  in the interval  $[x_{k-1}, x_k)$ . This  $x'$  is computable, based on Proposition 1.2.2. The above formula holds for all  $x$  for which  $T^k f(x)$  is defined for all  $k < 2^m$ .

The function  $S_{2^m} f$  can graphically be represented as



Define  $A_n f = nS_n f$ . Because  $T^{2^m} f = f$ ,

$$S_{i2^m} f = \frac{A_{i2^m}}{i2^m} = \frac{iA_{2^m}}{i2^m} = S_{2^m}.$$

Finally, for arbitrary  $n$ , since for some  $i$ ,  $i2^m \leq n < (i+1)2^m$ ,

$$\begin{aligned} \|S_n f - S_{2^m} f\| &= \|S_n f - S_{i2^m} f\| = \left\| \frac{A_n f}{n} - \frac{A_{i2^m} f}{i2^m} \right\| \\ &= \frac{\|i2^m(A_{i2^m} f + R_n) - nA_{i2^m} f\|}{ni2^m}, \end{aligned}$$

where  $R_n = A_n - A_{i2^m}$ .

This is in turn equal to

$$\left\| \frac{A_{i2^m} f(i2^m - n)}{ni2^m} + \frac{R_n}{n} \right\| \leq \left| \frac{i2^m - n}{n} \right| \|S_{2^m} f\| + \frac{\|R_n\|}{n}.$$

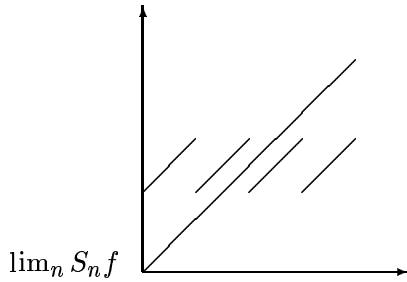
Since  $\|S_{2^m}\| \leq \|f\| < 1$ ,  $\left| \frac{i2^m - n}{n} \right| < \frac{2^m}{i2^m} = \frac{1}{i}$ , and

$$\|R_n\| = \|T^{i2^m} f + \dots + T^{n-1} f\| \leq (n - i2^m)\|f\| < 2^m,$$

it follows that

$$\|S_n f - S_{2^m} f\| < \frac{2}{i}.$$

Thus, if  $a_k = 0$ ,  $\lim_n S_n f = f$ . If  $a_k \neq 0$ , then  $\lim_n S_n f = S_{2^m} f$ . Once again, graphically:



It remains to find a formula that will distinguish between the two limit functions, and consequently between the cases when  $a_k = 0$  and  $a_k \neq 0$ . We cannot directly compare integrals of limit functions, since the pointwise ergodic theorem states that the integral of the limit is equal to the integral of the original function. Instead, we can consider the restriction of  $\lim_n S_n f$  on the second half of each subinterval. It is clear from the above graph that the two have different measures. A brief computation shows this.

Let  $I_k = [x_k + 2^{-(k+2)}, x_{k+1})$  and  $\hat{f}_k = \lim_n S_n f \upharpoonright I_k = \lim_n S_n f \chi_{I_k}$ . In general, (WWKL) is needed to prove that a product of an integrable function and a characteristic function is integrable. In this case, however, it can be shown directly that this is the case:  $\hat{f}_k$  is a piecewise continuous function and its integrability is shown using Lemma 4.5.2.

If  $a_k = 0$ , then  $\mu(\hat{f}_k) = 2^{-(k+2)}(1 - 2^{-(k+1)} - 2^{-(k+2)} + 2^{-(k+3)})$ .

If  $a_k \neq 0$ , then  $\mu(\hat{f}_k) = 2^{-(k+2)}(1 - 2^{-(k+1)} - 2^{-(k+2)})$ .

Therefore

$$a_k \neq 0 \leftrightarrow \mu(\hat{f}_k) = 2^{-(k+2)}(1 - 2^{-(k+1)} - 2^{-(k+2)}),$$

and the latter formula is  $\Pi_1$ , which completes the proof. □

It can be shown using the same argument that the  $L_1$  ergodic theorem (statement that the sequence  $\langle S_n f \rangle$  converges in the  $L_1$  norm for all  $f$ ) also reverses to (ACA). The transformation  $T$  remains the same, as does the proof. This is because we never used any of the pointwise properties of the limit function in the above theorem.



## Chapter 6

# Closing Remarks

It is my hope that the reader has been convinced that doing analysis in second-order arithmetic is both a fascinating and useful endeavor, that uncovers relationships that are easily overlooked in the standard mathematical practice.

This is by no means an attempt to formalize *all* of measure theory and *all* of functional analysis within second-order arithmetic. My purpose was to give a self-contained account of the ergodic theorems and results required to prove them. Much more work can and should be done towards a more complete understanding of these spaces within second-order arithmetic. A natural continuation of the work presented in Chapters 2 and 3, and in more breadth with Jeremy Avigad in [2], would include the analysis of the spectral theory of Hilbert spaces, as well as consideration of more general classes of Banach spaces. As for measure theory, there is the question of whether there are situations in which it is necessary to introduce measurable functions, and if so, how. It is also still unclear if it is possible to create a satisfactory theory of measure starting from sets. Finally, it would be especially worthwhile to explore and fully grasp the relationship between reverse and constructive mathematics.

Despite of my effort to make this presentation complete, a number of questions still remain open. Those pertaining to Hilbert spaces and the mean ergodic theorem were discussed in [2]. The list given below can be found in Section 16 of that paper.

1. Every closed linear subset of a Banach space is located.
2. Every closed linear subset of a Hilbert space is located.
3. Every closed linear subset of a Banach space is a closed subspace.

4. Every closed linear subset of a Hilbert space is a closed subspace.
5. If  $T$  is any bounded linear operator from a Banach space to itself and  $\lambda$  any real number,  $\{x \mid Tx = \lambda x\}$  is a closed subspace.
6. If  $T$  is any bounded linear operator from a Banach space to itself,  $\{x \mid Tx = x\}$  is a closed subspace.
7. If  $T$  is any bounded linear operator from a Hilbert space to itself,  $\{x \mid Tx = x\}$  is a closed subspace.

In  $\text{RCA}_0$ , all of these are implied by  $(\Pi_1^1\text{-CA})$ , and all, in turn, apply  $(ACA)$ . In addition, 1 implies all the statements below it; 2 implies 4 and 7; 3 implies all the statements below it; 4 implies 7; 5 is equivalent to 6 (since if  $\lambda \neq 0$ , we can define  $T'x = \frac{1}{\lambda}Tx$ ), and these in turn imply 7. It is possible, however, that all the statements are equivalent to  $(\Pi_1^1\text{-CA})$ , and it is also possible that they are all equivalent to  $(ACA)$ .

Also left wide open is the strength of the statement:

- If  $M$  is any closed linear subset of a Hilbert space, and  $x$  is any point, and the distance from  $x$  to  $M$  exists, then the projection of  $x$  on  $M$  exists.

There are also many unresolved questions in the realm of measure theory. I proved a number of results under the condition that the functions considered are essentially bounded, but in some instances was not able to conclude anything about the general case. It is especially unfortunate that the definitions of  $f \cdot f$  and  $f^2$  in general cannot be shown to be equivalent in  $\text{RCA}_0$ . Also, I tailored the definition of products and powers of  $L_p$  functions to suit my particular needs in this work; the definition may be too restrictive and may not recognize all functions that are classically integrable as such. This should also be addressed in the future.

Another unresolved issue is this: it is a known fact that  $(WWKL)$  proves that the domain of an integrable function is a full set. What is not known, however, is whether this is sharp, that is, if that statement reverses to weak-weak König's lemma.

It would also be quite interesting to figure out the nuances regarding the exact relationship between measure theory in Bishop-style mathematics and in our framework.



# Appendix A

To pay off a debt from the introduction, we show that the statement “ $C(X)$  is an integration space” (in the classical sense) reverses to (*WWKL*). The constructive definition of an integration space is given in [8] and is as follows:

**Definition A.0.2** *An integration space is a triple  $(X, L, I)$  with the following properties:*

- i) If  $f, g, \in L$  and  $\alpha, \beta \in \mathbb{R}$ , then  $\alpha f + \beta g$ ,  $|f|$  and  $f \wedge 1$  belong to  $L$  and  $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ .*
- ii) If  $f \in L$  and  $\langle f_n \rangle$  is a sequence of nonnegative functions in  $L$  such that  $\sum_n I(f_n)$  converges and  $\sum_n I(f_n) < I(f)$ , then there exists  $x$  in  $X$  such that  $\sum_n f_n(x)$  converges and  $\sum_n f_n(x) < f(x)$ .*
- iii) There exists a function  $p$  in  $L$  with  $I(p) = 1$ .*
- iv) For each  $f$  in  $L$ ,  $\lim_n I(f \wedge n) = I(f)$  and  $\lim_n I(|f| \wedge n^{-1}) = 0$ .*

The classical definition of an integration space in the Daniell integral theory is almost the same as the one above, and can be found, for example, in [24], where Daniell integrals are discussed at length. The essential difference lies in item ii). In the standard definition, it is replaced with the continuity condition (A):

Let  $\langle f_n \rangle$  be a sequence of functions in  $L$  such that  $f_n(x) \downarrow 0$  for all  $x$ . Then  $I(f_n) \downarrow 0$ .

Classically, the two are equivalent, but constructively, ii) is stronger.

We will show that the statement “(A) holds in every space  $C(X)$  induced by a complete separable metric space  $X$ ” implies (*WWKL*). In fact, the space that provides this reversal is  $C([0, 1])$ , so it may be more accurate to say that the statement “(A) holds in  $L = C([0, 1])$ ” implies (*WWKL*). The proof will be based on the following result [26, 32, 11]:

**Theorem A.0.3** *The following assertions are equivalent over  $\text{RCA}_0$ :*

1. (WWKL).
2. *For any covering of the closed unit interval  $[0, 1]$  by a sequence of open intervals  $\langle (a_n, b_n) \mid n \in \mathbb{N} \rangle$ , we have  $\sum_n |b_n - a_n| \geq 1$ .*

We will show that (A) implies (2) above, or rather that  $\neg(2) \rightarrow \neg(A)$ . If (2) does not hold, there exists a covering of  $[0, 1]$  by  $\langle (a_n, b_n) \rangle$  such that  $\sum_n |b_n - a_n| < 1$ .

Now consider  $C([0, 1])$  and construct a sequence of functions  $\langle f_n \rangle$  that fails (A), using the covering above. Let  $f_n$  be 1 on  $[a_i + \frac{b_i - a_i}{2^{n+2}}, b_i - \frac{b_i - a_i}{2^{n+2}}]$  for  $i \leq n$ , 0 on  $[0, 1] \setminus \bigcup_{i \leq n} [a_i, b_i]$  and linear everywhere else. This is an increasing sequence of functions in  $C([0, 1])$ . Furthermore,  $\lim_n f_n(x) = 1$  for all  $x$ , because  $\langle (a_n, b_n) \rangle$  is a covering. At the same time,  $\mu(f_n) \leq \sum_{i \leq n} |b_i - a_i|$ . Because  $\sum_n |b_n - a_n| < 1$ , if  $\lim_n \mu(f_n)$  exists, it is less than 1. In other words,  $f_n(x) \uparrow 1$  for all  $x$ , but  $\mu(f_n) \not\rightarrow \mu(1)$ . This immediately implies  $\neg(A)$  and so the classical definition of an integration space implies the weak-weak König's lemma.

However, this conclusion no longer holds if we replace the classical definition of integration spaces with that by Bishop and Bridges. It cannot be proved in  $\text{RCA}_0$  that item ii) in Definition A.0.2 and property (A) are equivalent.

To prove this fact, suppose (A), and let  $\langle f_n \rangle$  be a sequence satisfying the hypotheses of ii). Define a sequence  $\langle g_i \rangle$  such that

$$\begin{aligned} g_0 &= f \\ g_{i+1} &= g_i - f_i. \end{aligned}$$

Then for each  $n$ ,  $g_n = f - \sum_{i < n} f_i$ , and the hypothesis implies that  $\mu(g_n)$  is decreasing and converges, but not to 0. Applying (A), we conclude that for some  $x$ ,  $g_n(x)$  does not converge to 0. In other words, for some  $x$ , either  $\sum_n f_n(x)$  does not converge at all, or it converges to something other than  $f(x)$ . But this is different from conclusion of ii), which does not allow the first possibility.

This last argument makes it clear that Bishop and Bridges were very careful in choosing definitions that go through constructively.

## Appendix B

A number of proofs in Chapter 4 used properties of derivatives of power functions. Instead of developing a theory of differentiation, this Appendix is intended only to convince the reader that power functions are well-defined, and that it is legitimate to differentiate them, that basic laws of differentiation hold, along with the mean value theorem and some other simple results. These are all concepts that are usually taken for granted, but require some effort in reverse and constructive mathematics. The motivation comes from Bishop's work [6, 8], and, more importantly, from Schwichtenberg, who worked out most of these issues constructively in [25]. Most technical details of proofs will be omitted. Assume throughout that the underlying system is  $\text{RCA}_0$ .

The first question to answer is: how to define the function  $x^p$ ? The answer is clear for  $p \in \mathbb{N}$ , but more difficult when  $p$  is rational, and even more so when it is an irrational number. All standard calculus and analysis textbooks define power functions via exponentiation, i.e.  $x^p = e^{p \ln x}$  for all  $x > 0$ . This is the approach that we will also take, but prior to that we need to justify the validity of taking these operations, that is, we need to show that  $e^x$  and  $\ln x$  (when  $x > 0$ ) are continuous.

Based on Lemma II.6.5. in [26], power series give rise to continuous functions, so  $e^x$  can be defined as the sum of the series  $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ . Schwichtenberg shows in [25] that  $e^x \cdot e^y = e^{x+y}$ . He then proceeds to define  $\ln x = \int_1^x \frac{1}{t} dt$ , for  $x > 0$ . It is not obvious that this integration is permitted in  $\text{RCA}_0$ . In fact, according to Lemma IV.2.6 in [26],  $\frac{1}{x}$  can be integrated only if it has a modulus of uniform continuity, but it is a well-known fact that this is true, as long as we restrict ourselves to a closed interval  $[a, b]$ , with  $a > 0$ .

A function obtained via integration is continuous and differentiable (definition of derivatives will follow below), hence  $\ln x$  is continuous for all  $x > 0$ . Furthermore, the reader can find a proof in [25] that  $\ln x$  thus defined has the usual properties and that  $e^x$  and  $\ln x$  are inverse to each other.

This means that when  $x > 0$ , the function  $f(x) = e^{p \ln x}$  as a composition of continuous functions is itself continuous ([26], Lemma II.6.4) and we let  $x^p = e^{p \ln x}$ . Standard properties of power functions can be shown, for example that  $x^{p+q} = x^p \cdot x^q$  and  $(x^p)^q = x^{pq}$  (and in particular,  $(x^p)^{1/p} = x$ ). Also, if  $p \in \mathbb{N}$ ,  $x^p = \underbrace{x \cdot \cdots \cdot x}_p$ .

There is only one problem with this definition, and that is that it only works for  $x > 0$ . However, it is possible to show that  $\lim_{x \rightarrow 0} e^{p \ln x} = 0$ , by showing that  $\ln x$  is negative and unbounded when  $x \rightarrow 0$  and that  $e^x$  is unbounded when  $x \rightarrow \infty$ . We also know that  $x^p$  is a continuous function for all  $x > 0$ , which means that there is a sequence of quintuples (as described in Definition 1.1.4) that codes it. With some care, it can be shown that, if we add to this sequence a countable number of conditions specifying that  $0 \mapsto 0$ , we will obtain a code for another continuous function, the function  $x^p$  for all nonnegative numbers.

Another useful fact can be shown: suppose  $p$  is an arbitrary real number, hence represented as the limit of a strong Cauchy sequence. Then it follows from properties of exponents and logarithms that  $x^p = \lim_n x^{p_n}$  when  $x > 0$  (the fact is immediate when  $x = 0$ ).

We can now define  $f^p$ , when  $f$  is a nonnegative simple function. With some effort it can be shown that if  $f \in S(X)$ , then  $f^p \in C(X)$ , by showing it has a modulus of uniform continuity.

Next we need to discuss some basic concepts regarding differentiation. The definition of derivative is what one would expect. A function  $f$  is differentiable at  $x$  if  $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$  exists. This limit, if it exists, is  $f'(x)$ . Since the goal is to formalize all this in  $\text{RCA}_0$ , the rate of convergence has to be computable. Bishop gives the following definition:

**Definition B.0.4** *Let  $f$  and  $g$  be continuous functions on a compact proper interval  $I$  such that for each  $\varepsilon$  there exists  $\delta(\varepsilon) > 0$  with*

$$|f(y) - f(x) - g(x)(y - x)| \leq \varepsilon |y - x|$$

*whenever  $x, y \in I$  and  $|y - x| \leq \delta(\varepsilon)$ .*

and with it proceeds to prove the following:

**Proposition B.0.5** *With the notation  $Df = f' = \frac{df}{dx}$ , the following hold:*

1.  $D(f_1 + f_2) = Df_1 + Df_2$ .
2.  $D(f_1 f_2) = Df_1 f_2 + f_1 Df_2$ .

3.  $D(f_1^{-1}) = -f_1^{-2}Df_1$ .
4.  $\frac{dx}{dx} = 1$ .
5.  $\frac{dc}{dx} = 0$ .
6.  $(g \circ f)' = (g' \circ f)f'$ .

The standard proofs of these facts are constructive: they apply almost word for word in  $\text{RCA}_0$ .

The statements about power functions that were mentioned in the main text and require proper proofs are these:

**Lemma B.0.6** ( $\text{RCA}_0$ ) 1. *The inequality  $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$  holds for all non-negative real numbers  $a$  and  $b$ . (4.3.1)*

2. *The inequality  $|x - y|^p \leq |x^p - y^p|$  holds for all  $x$  and  $y$ . (4.3.2)*

3. *The inequality  $|x^p - y^p| \leq p|(x - y)(x^{p-1} + y^{p-1})|$  is true for all real  $x$  and  $y$ . (4.3.2)*

4. *For all  $x, y \in \mathbb{R}^+$ , and  $q > p$ ,  $y^{q/p} \geq \frac{q}{p} x^{\frac{q-p}{p}}(y - x) + x^{q/p}$ . (4.3.3)*

To prove Lemma B.0.6, a number of facts are needed.

**Proposition B.0.7** *The function  $f(x) = x^p$  is differentiable for all real numbers  $p$  and  $f'(x) = px^{p-1}$ .*

*Proof.* If  $p$  is a natural number, this is proved by induction, using the product rule. The base case is item 4. in Proposition B.0.5. If  $p$  is rational, the result is provided by the chain rule. Finally, if  $p$  is an arbitrary real number,

$$(x^p)' = (e^{p \ln x})' = \frac{p}{x} e^{p \ln x} = px^{p-1},$$

by another application of the chain rule and by properties of exponents and logarithms.  $\square$

**Proposition B.0.8** *The function  $f(x) = x^p$  is convex for all  $p > 1$ , that is, it is above its tangent line at every point.*

*Proof.* The standard proof that if  $f'' > 0$ , then  $f$  is convex, applies. It is easy to see that when  $p > 1$ ,  $(x^p)'' = p(p-1)x^{p-2} > 0$ .  $\square$

**Proposition B.0.9** *The mean value theorem for  $x^a$ : for every  $x$  and  $y$  there is a  $\xi$  between them such that  $\frac{x^a - y^a}{x - y} = a\xi^{a-1}$ .*

*Proof.* The general form of the mean value theorem was proved by Hardin and Velleman in [17]. The proof of this theorem in  $\text{RCA}_0$  is not at all obvious.  $\square$

**Proposition B.0.10** *If  $f$  is differentiable at  $x$  and  $f'(x) > 0$ , then  $f$  is increasing in some neighborhood of  $x$ ; if  $f'(x) < 0$ ,  $f$  is decreasing; consequently, if  $f'(x_0) = 0$ ,  $f'(x) < 0$  (resp.  $f'(x) > 0$ ) for all  $x < x_0$  and  $f'(x) > 0$  (resp.  $f'(x) < 0$ ) for all  $x > x_0$ , then  $f(x_0)$  is the absolute minimum (resp. absolute maximum) of  $f$ .*

*Proof.* If  $f'(x) > 0$ , then based on properties of limits and continuous functions, there is some neighborhood of  $x$  for which  $\frac{f(x) - f(y)}{x - y} > 0$ . This means that for  $y < x$  in that neighborhood,  $f(x) > f(y)$  and for  $y > x$ ,  $f(x) < f(y)$  as required. Other facts are shown similarly.  $\square$

We can now outline the proof of Proposition B.0.6:

*Proof.*

1. The proof consists of finding the maximum of the function  $b \mapsto ab - \frac{a^p}{p} - \frac{b^q}{q}$ . Since  $f'(x) = a - x^{q-1}$ , and  $f'(x) > 0$  when  $x > a^{p/q}$ , while  $f'(x) < 0$  when  $x < a^{p/q}$ , the maximum is attained when  $x = a^{\frac{1}{q-1}} = a^{p/q}$  and is equal to  $a^{1+p/q} - \frac{a^p}{p} - \frac{a^p}{q} = 0$ . Therefore,  $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ .
2. Let  $x \geq y \geq 0$  and divide the entire inequality by  $x^p$ . Then it suffices to show  $(1 - t)^p \leq 1 - t^p$  for  $0 \leq t \leq 1$ . As in the previous item, we consider the derivative. In this case,  $f(t) = (1 - t)^p - 1 + t^p$ , while  $f'(t) = -p(1 - t)^{p-1} + t^{p-1}$ , so  $f'(t) > 0$  when  $t < \frac{1}{2}$  and  $f'(t) < 0$  when  $t > \frac{1}{2}$ . Clearly, the minimum of the function on the interval is attained at  $t = 0$  or  $t = 1$ . Since  $f(0) = f(1) = 0$ ,  $f(t) \geq 0$  which implies the original claim.
3. This inequality is proved using the mean value theorem. If  $x, y \geq 0$ , (at least one is nonzero, otherwise the claim is trivial),

$$\frac{x^p - y^p}{x - y} = p\xi^{p-1} \leq p(x^{p-1} + y^{p-1}).$$

4. Let  $f(x) = x^{q/p}$ . Then the inequality can be written as  $\frac{f(y) - f(x)}{y - x} \geq f'(x)$ , which is true since  $f$  is a convex function, as  $\frac{q}{p} > 1$ .  $\square$

# Bibliography

- [1] Jeremy Avigad and Erich Reck. “Clarifying the nature of the infinite”: the development of metamathematics and proof theory. Technical Report CMU-PHIL-120, Carnegie Mellon University, 2001.
- [2] Jeremy Avigad and Ksenija Simic. Fundamental notions of analysis in subsystems of second-order arithmetic. Draft available at <http://www.andrew.cmu.edu/~avigad/>, 2003.
- [3] Patrick Billingsley. *Ergodic Theory and Information*. Wiley, New York, 1965.
- [4] G.D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17:656–660, 1931.
- [5] Errett Bishop. An upcrossing inequality with applications. *Michigan Math. J.*, 13:1–13, 1966.
- [6] Errett Bishop. *Foundations of Constructive Analysis*. McGraw-Hill, New York, 1967.
- [7] Errett Bishop. A constructive ergodic theorem. *J. Math. Mech.*, 17:631–639, 1967/1968.
- [8] Errett Bishop and Douglas Bridges. *Constructive Mathematics*. Springer, Berlin, 1985.
- [9] Douglas Bridges. *Constructive functional analysis*. Research Notes in Mathematics. Pitman Publishing Limited, 1979.
- [10] Douglas K. Brown. Notions of closed subsets of a complete separable metric space in weak subsystems of second-order arithmetic. In *Logic and computation (Pittsburgh, PA, 1987)*, pages 39–50. Amer. Math. Soc., Providence, RI, 1990.

- [11] Douglas K. Brown, Mariagnese Giusto, and Stephen G. Simpson. Vitali's theorem and WWKL. *Arch. Math. Logic*, 41(2):191–206, 2002.
- [12] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. Spectral operators, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1971 original, A Wiley-Interscience Publication.
- [13] Matthew Frank. Ergodic theory in the 1930s: a study in international mathematical activity.  
Draft available at <http://math.uchicago.edu/~mfrank/erg.pdf>.
- [14] Harvey Friedman, Stephen G. Simpson, and Xiaokang Yu. Periodic points and subsystems of second-order arithmetic. *Ann. Pure Appl. Logic*, 62(1):51–64, 1993. Logic Colloquium '89 (Berlin).
- [15] A.M. Garsia. A simple proof of Eberhard Hopf's maximal ergodic theorem. *J. Math. Mech.*, 14:381–382, 1965.
- [16] Paul R. Halmos. *Lectures on ergodic theory*. Chelsea Publishing Co., New York, 1960.
- [17] Christopher S. Hardin and Daniel J. Velleman. The mean value theorem in second order arithmetic. *J. Symbolic Logic*, 66(3):1353–1358, 2001.
- [18] Teturo Kamae. A simple proof of the ergodic theorem using nonstandard analysis. *Israel J. Math.*, 42(4):284–290, 1982.
- [19] Yitzhak Katznelson and Benjamin Weiss. A simple proof of some ergodic theorems. *Israel J. Math.*, 42(4):291–296, 1982.
- [20] J. A. Nuber. A constructive ergodic theorem. *Transactions of the American Mathematical Society*, 164:115–137, 1972.
- [21] Karl Petersen. *Ergodic Theory*, volume 2 of *Cambridge studies in advanced mathematics*. Cambridge University Press, Cambridge, 1983.
- [22] Marian B. Pour-El and J. Ian Richards. *Computability in analysis and physics*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1989.
- [23] Frédéric Riesz. Sur la théorie ergodique. *Comment. Math. Helv.*, 17:221–239, 1945.



- [24] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.
- [25] Helmut Schwichtenberg. Constructive analysis with witnesses. Draft available at <http://www.mathematik.uni-muenchen.de/~schwicht/>.
- [26] Stephen G. Simpson. *Subsystems of Second-Order Arithmetic*. Springer, Berlin, 1998.
- [27] Bas Spitters. *Constructive and intuitionistic integration theory and functional analysis*. PhD thesis, University of Nijmegen, 2002.
- [28] John von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Nat. Acad. Sci.*, 18:70–82, 1932.
- [29] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.
- [30] Xiaokang Yu. Riesz representation theorem, Borel measures and subsystems of second-order arithmetic. *Ann. Pure Appl. Logic*, 59(1):65–78, 1993.
- [31] Xiaokang Yu. Lebesgue convergence theorems and reverse mathematics. *Math. Logic Quart.*, 40(1):1–13, 1994.
- [32] Xiaokang Yu and Stephen G. Simpson. Measure theory and weak König's lemma. *Arch. Math. Logic*, 30(3):171–180, 1990.