

2004

Fundamental Notions of Analysis in Subsystems of Second-Order Arithmetic

Jeremy Avigad

Carnegie Mellon University, avigad@cmu.edu

Ksenija Simic

University of Arizona - Tucson

Follow this and additional works at: <http://repository.cmu.edu/philosophy>

 Part of the [Philosophy Commons](#)

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Philosophy by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Fundamental notions of analysis in subsystems of second-order arithmetic

Jeremy Avigad

*Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213*

Ksenija Simic

*Department of Mathematics
The University of Arizona
Tucson, AZ 85721*

Abstract

We develop fundamental aspects of the theory of metric, Hilbert, and Banach spaces in the context of subsystems of second-order arithmetic. In particular, we explore issues having to do with distances, closed subsets and subspaces, closures, bases, norms, and projections. We pay close attention to variations that arise when formalizing definitions and theorems, and study the relationships between them. For example, we show that a natural formalization of the mean ergodic theorem can be proved in ACA_0 ; but even recognizing the theorem's "equivalent" existence assertions as such can also require the full strength of ACA_0 .

Contents

1	Introduction	2
2	Preliminaries	5
3	Leftmost paths through trees	7
4	Complete separable metric spaces	9
5	Closed sets	11

Email addresses: avigad@cmu.edu (Jeremy Avigad),
ksimic@math.arizona.edu (Ksenija Simic).

URLs: <http://www.andrew.cmu.edu/user/avigad> (Jeremy Avigad),
<http://math.arizona.edu/~ksimic/> (Ksenija Simic).

6	Distances from a point to a set	14
7	Closures	17
8	Iterative functions on metric spaces	19
9	Hilbert spaces and Banach spaces	21
10	Bases and independent generating sequences	23
11	Closed subspaces	30
12	Distances and projections	32
13	Norms and kernels of bounded linear functionals	37
14	The mean ergodic theorem	41
15	Complex Hilbert spaces and the remaining reversals	46
16	Final remarks	54
	References	55

1 Introduction

A good deal of work in the foundations of mathematics in the nineteenth century was directed towards grounding mathematics, and analysis in particular, in the theory of the natural numbers. Of course, constructions of the real numbers, like those of Weierstrass, Dedekind, and Cantor, required an appropriate “logical,” or set-theoretic background; and almost as soon as there were full-blown set-theoretic foundations for mathematics, there were concerted efforts to determine the extent to which portions of analysis can be carried out in restricted fragments. Weyl’s landmark *Das Kontinuum* [26], Hilbert and Bernays’ *Grundlagen der Mathematik* [10], and, more recently, Takeuti’s *Two Applications of Logic to Mathematics* [24] contributed to this program.

Such formalizations of analysis are often couched in terms of restricted subsystems of second-order arithmetic. In this context, the “reverse mathematics” program, promoted principally by Harvey Friedman and Stephen Simpson, aims to calibrate the strength of central mathematical theorems in terms of the axiomatic set existence principles that are needed to prove them.

Because set-theoretic language and terminology pervade modern mathematics, the first step in the reverse mathematician’s analysis is to adapt the rel-

evant definitions and concepts to the language of second-order arithmetic. At the outset, this can pose problems, since infinitary mathematical objects and structures must ultimately be represented in terms of sets of natural numbers. This can force one to restrict one’s attention, say, to countable algebraic structures and separable spaces. In the language of second-order arithmetic, there is, for example, no way to represent the notion of an arbitrary function between separable metric spaces; one may reasonably restrict one’s attention to continuous functions, but then these have to be defined indirectly, in terms of their countable representations, rather than as arbitrary functions that happen to satisfy the usual epsilon-delta characterization.

To make matters worse, the body of theorems that one can ultimately derive in restricted theories can be sensitive to one’s choice of definitions. For example, Cauchy sequences are often easier to deal with than Dedekind cuts of rationals, and even with the choice of the former it makes a difference whether or not one requires an explicit rate of convergence. Other examples abound; for example, it takes some axiomatic strength to prove the equivalence of various formulations of compactness for separable metric spaces (see [2,22]), as well as the equivalence of various notions of a closed subset (see [3]). It can also take some axiomatic strength to obtain moduli of uniform continuity for continuous functions on a compact space (see [22,14]), or distance functions for closed sets (see [8,7]). 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 prologue to Bishop and Bridges’ *Constructive Mathematics* [1], 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. This is certainly the case, for example, with respect to finitary notions from number theory and combinatorics. 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. When it comes to developing a mathematical theory in a restricted framework,

- various natural definitions may turn out to be provably equivalent in a weak base theory;
- among definitions that are not provably equivalent, one may prove to be more natural, or more useful; or

- different definitions may prove to be useful in different contexts.

We maintain that in each case, something interesting has been learned. Thus one can view reverse mathematics as a study of the ways in which mathematical theories can (or have to) be developed if one is committed to avoiding abstract set-theoretic notions in favor of more explicit representations. Such a study can be interesting in its own right, for the mathematical insights it brings, or for the mathematical questions it raises; it can also be of use to logical analyses which aim to extract algorithms and other useful information from classical mathematical developments.

Our goal here is to present an analysis of a number of fundamental notions from the theories of metric, Hilbert, and Banach spaces in the context of subsystems of second-order arithmetic. We believe that the results presented here support the contentions above. For example, in Section 5, we observe that there are at least three fundamentally different notions of a closed subset of a complete separable metric space; and in Section 11 we note that these lift to three different notions of a closed subspace of a Hilbert space or a Banach space. We invest a good deal of effort in clarifying the relationships between these notions, and understanding situations in which they arise. We will see, for example, that the distinctions are important in an analysis of von Neumann's mean ergodic theorem, stated in the general context of a Hilbert space; and, in the other direction, our analysis of the mean ergodic theorem turns out to be quite helpful in sorting out the relationships between the various notions.

The formalization of mathematics in subsystems of second-order arithmetic is closely related to work carried out in the fields of constructive mathematics and recursive mathematics. There are key differences between reverse mathematics and these other fields, however. In one sense, our work is less restricted, since we allow the use of classical reasoning and noncomputable constructions. Indeed, our goal is often to clarify the extent to which non-computable constructions are necessary. In another sense, however, we are more constrained, in that we pay careful attention to the axiomatic framework in which the constructions take place. This attention makes it possible to subject the formal developments to proof-theoretic analysis, as in, say, Kohlenbach's "proof mining" program [16,15].

In any event, in the present work we have drawn on ideas from the literature in all three subjects. In particular, we have benefited a good deal from the constructive developments of the theory of metric, Hilbert, and Banach spaces, especially those of Bishop and Bridges [1], and Spitters [23]; from the recursive development of Hilbert and Banach space theory in Pour-El and Richards [20]; and, of course, from a number of works in reverse mathematics, including [22,3,6,8]. We are especially grateful to the anonymous referee for a very careful reading, and numerous suggestions and corrections.

2 Preliminaries

We will assume familiarity with the study of subsystems of second-order arithmetic as in [22]. To recap the essentials: the language of second-order arithmetic is a two-sorted language with variables x, y, z, \dots intended to denote natural numbers, and X, Y, Z, \dots intended to denote sets of numbers. The language has $0, 1, +, \times$, and $<$, as well as a binary relation \in to relate the two sorts. The notations $\Sigma_n^0, \Pi_n^0, \Sigma_n^1$, and Π_n^1 denote the usual syntactic hierarchies, and we will drop the superscripted 0 in the first two. All the theories we will consider here include the schema of induction for Σ_1 formulas, which are allowed to have number and set parameters; in particular, we always have the induction axiom for any set, i.e. induction for the formula $x \in X$. What distinguishes the theories from one another are their set existence principles. The base theory, RCA_0 , is based on the schema of *recursive comprehension axioms*, (RCA):

$$\forall x (\varphi(x) \leftrightarrow \psi(x)) \rightarrow \exists Z \forall x (x \in Z \leftrightarrow \varphi(x)),$$

where φ and ψ are Σ_1 and Π_1 respectively. Intuitively, this asserts that if a class of numbers is both a computably enumerable and co-computably enumerable (and so, computable), this class forms a set. Similarly, the theory ACA_0 is based on the comprehension schema for arithmetic formulas, denoted (ACA), and the theory $\Pi_1^1\text{-CA}_0$ is based on the comprehension schema for arithmetic Π_1^1 formulas, denoted ($\Pi_1^1\text{-CA}$). Of the two remaining theories in the main reverse mathematics hierarchy, WKL_0 will play only a small role here, and ATR_0 will not come up at all.

Working in RCA_0 , one can code various finite objects like rational numbers, pairs of natural numbers, or finite sequences of natural numbers, as numbers; and one can code e.g. functions from \mathbb{N} to \mathbb{N} and real numbers as sets of numbers. A pair or a countable sequence of sets of numbers can also be coded as a set of numbers. We will take such codings and their properties for granted; see [22] for details.

In RCA_0 , the arithmetic comprehension schema, (ACA), is equivalent to a single axiom that asserts that for each set Z , the Turing jump of Z exists. Here the Turing jump of Z can be taken to be the set $\{x \mid \exists y \theta(x, y, Z)\}$, where θ is Δ_0 and $\exists y \theta(x, y, Z)$ is a complete Σ_1 formula. The following lemma provides alternative characterizations that will be useful.

Lemma 2.1 (RCA_0) *Each of the following statements is equivalent to (ACA):*

- (1) *Every increasing sequence $\langle a_n \mid n \in \mathbb{N} \rangle$ of real numbers in $[0, 1]$ has a limit.*
- (2) *If $\langle b_n \mid n \in \mathbb{N} \rangle$ is any sequence of reals such that for each n , $\sum_{i < n} b_i^2 \leq 1$,*

then $\sum b_i^2$ exists.

(3) If $\langle c_n \mid n \in \mathbb{N} \rangle$ is any sequence of real numbers, the set $C = \{i \in \mathbb{N} \mid c_i \neq 0\}$ exists.

Proof. The equivalence of 1 with (ACA) can be found in [22]. Taking $a_n = \sum_{i < n} b_i^2$ shows 1 implies 2. Conversely, given $\langle a_n \rangle$ as in 1, let $b_0 = a_0$ and $b_n = \sqrt{a_n - a_{n-1}}$ for $n > 1$; then $\sum_{i < n} b_i^2 = a_n$ and $\sum b_n^2 = \lim_n a_n$.

Finally, the set C in 3 is easily obtained using arithmetic comprehension from the sequence $\langle c_n \rangle$. Conversely, given θ defining the Turing jump of Z as above, one can prove in RCA_0 that the sums

$$c_i = \sum_{\{j \mid \theta(i,j,Z)\}} 2^{-j},$$

exist, and clearly $c_i \neq 0 \leftrightarrow \exists j \theta(i, j, Z)$. \square

The next useful lemma is a formalization of the fact that any multifunction with a computably enumerable graph can be uniformized by a computable function.

Lemma 2.2 *Let $\varphi(x, y)$ be any Σ_1 formula, possibly with set and number parameters other than the ones shown. Then RCA_0 proves*

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

Proof. Using pairing we can assume that $\varphi(x, y)$ is of the form $\exists z \psi(x, y, z)$, where ψ is Δ_0 . Using (RCA), define $g(x)$ to be the least number w coding a pair $\langle y, z \rangle$ satisfying $\psi(x, y, z)$. (Clearly the graph of g has a Σ_1 description; but then the equivalence $g(x) \neq w \leftrightarrow \exists w' (g(x) = w' \wedge w \neq w')$ yields a Σ_1 description of the complement of the graph of g .) Using (RCA) again, let $f(x) = (g(x))_0$. \square

Finally, we gather some principles that can be justified on the basis of Σ_1 induction, although they may initially appear to be considerably stronger. In the statement of the next lemma, the class of $\Delta_0(\Sigma_1)$ formulas is defined to be the smallest class of formulas containing the Σ_1 formulas and closed under boolean operations (including negation) as well as bounded quantification.

Lemma 2.3 *The following induction principles are available are derivable in RCA_0 :*

- (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.

One can prove Lemma 2.3 by showing that Σ_1 induction can be used to justify a comprehension principle for finite sets defined by $\Delta_0(\Sigma_1)$ formulas; see, for example, [22, Theorem II.3.9] or [9, Lemma 2.14].

3 Leftmost paths through trees

A *tree on* $\{0, 1\}$ is defined to be a set of finite sequences from $\{0, 1\}$, closed under initial segments. Similarly, a *tree on* ω is a set of finite sequences from \mathbb{N} closed under initial segments. In either case, a *path* through such a tree T is defined to be a function f such that every initial segment of f is in T . A *leftmost* path f through T is one that is least in the lexicographic ordering, so that if g is any other path through T , then $f(i) < g(i)$ for the least i at which f and g differ. In this section, we show that statements asserting the existence of leftmost paths through non-well-founded trees on $\{0, 1\}$ and ω are equivalent to (ACA) and $\Pi_1^1\text{-CA}$, respectively. These facts will be used in Section 6.

Lemma 3.1 (RCA_0) *(ACA) is equivalent to the assertion that any infinite tree on $\{0, 1\}$ has a leftmost path.*

Proof. To show using (ACA) that every infinite tree has a leftmost path, use arithmetic comprehension to define the set S of elements of T with infinitely many elements extending them. It is easy to define the leftmost path recursively from this set; see [22] for details. Conversely, suppose every tree has a leftmost path, and let us show that for every X the Turing jump of X exists. Given X and θ as above, put a binary sequence σ in T if and only if

$$\forall x < \text{length}(\sigma) (\exists y < \text{length}(\sigma) \theta(x, y, X) \rightarrow (\sigma)_x = 1).$$

In other words, whenever there is a witness less than the length of σ that x is in the Turing jump of X , $(\sigma)_x = 1$. Since every finite sequence of 1's is in T , T is infinite. If f is any path through T and $\exists y \theta(x, y, X)$, taking an initial segment of f long enough shows that $f(x) = 1$. Conversely, if f is a leftmost path, whenever $f(x) = 1$ it is the case that $\exists y \theta(x, y, X)$; otherwise, we could obtain a path further to the left by changing $f(x)$ to 0. \square

In RCA_0 , Π_1^1 comprehension is equivalent to Σ_1^1 comprehension. With (ACA) , one can use skolemization to show that every Σ_1^1 formula $\varphi(x, Z)$ is equivalent to one of the form

$$\exists f \forall \sigma \subset f \theta(x, \sigma, Z),$$

where f is a function from \mathbb{N} to \mathbb{N} , $\sigma \subset f$ means that σ is a finite initial segment of f , and θ is Δ_0 . Modifying θ we can assume that for each x the

set of σ satisfying $\theta(x, \sigma, Z)$ is closed under initial segments. This leads to the following lemma (details can be found in [22]):

Lemma 3.2 (RCA_0) Π_1^1 comprehension is equivalent to the assertion that if T_x is any sequence of trees on ω , then there is a set S such that for x , x is in S if and only if there is a path through T_x .

The following alternative characterization of Π_1^1 comprehension will be useful to us below. It appears in [18, Theorem 6.5] with a slightly different proof; the recursion-theoretic analog may well be folklore, but we have been unable to find a reference.

Lemma 3.3 (RCA_0) Π_1^1 comprehension is equivalent to the assertion that for every tree T on ω , if T has a path, it has a leftmost path.

Proof. The leftmost path through T can be defined recursively in the set of nodes of T through which there is a path, and this set, in turn, is Σ_1^1 -definable relative to T . This construction can be formalized straightforwardly in $\Pi_1^1\text{-CA}_0$; see [22] for details.

Conversely, to prove Π_1^1 comprehension from the leftmost path principle, let $\varphi(x, Z)$ be the Σ_1^1 formula $\exists f \forall \sigma \subset f \theta(x, \sigma, Z)$ as above. The idea is to define a tree T , recursive in Z , such that a function g is a path through T if and only if it satisfies the following conditions:

- (1) $1 - g(0), 1 - g(2), 1 - g(4), \dots$ is the characteristic function of a set A ; in other words, for every x , $\chi_A(x) = 1 - g(2x)$.
- (2) This set A is a subset of $\{x \mid \varphi(x, Z)\}$. In other words, whenever x is in A , there is a function f_x witnessing $\forall \sigma \subset f_x \theta(x, \sigma, Z)$.
- (3) The sequence $g(1), g(3), g(5), \dots$ codes this information, in the sense that for each x in A , the function

$$f_x(y) = g(2(\langle x, y \rangle) + 1)$$

is such a witness.

We will assume that our pairing function $\langle x, y \rangle$ is monotone in x and y , and so, in particular, $\langle x, y \rangle \geq x$ for every x and y . Clearly the sequence $\lambda x 1$ (corresponding to $A = \emptyset$) meets the three criteria above. On the other hand, a leftmost path will have the property that as many even values of g as possible will be 0, which is to say, A will be as big as possible. This will imply $\{x \mid \varphi(x, Z)\}$ is a subset of A , and so equal to A .

The definition of T is as follows. To decide whether or not a finite sequence τ is in T , first write τ as $\langle a(0), b(0), \dots, a(k-1), b(k-1) \rangle$; if the length of τ is odd ignore the last element. Put τ in T if and only if for each $x < k$, $a(x)$ is

either 0 or 1, and every initial segment σ of

$$\langle b(\langle x, 0 \rangle), \dots, b(\langle x, y \rangle) \rangle$$

satisfies $\theta(x, \sigma, Z)$, where y is the largest number such that $\langle x, y \rangle < k$. Clearly any path through T satisfies the conditions above.

Suppose now that g is the leftmost path; we only need to show $\{x \mid \varphi(x, Z)\}$ is a subset of A . Suppose not; then there is an x and an f_x such that $g(2x) = 1$ (implying $x \notin A$) but f_x satisfies $\forall \sigma \subset f_x \theta(x, \sigma, Z)$. But then we can define g' by

$$\begin{aligned} g'(2x) &= 0 \\ g'(2\langle x, y \rangle + 1) &= f_x(y) \quad \text{for every } y \\ g'(z) &= g(z) \quad \text{for other values of } z \end{aligned}$$

Then g' is also a path through T , with $g'(2x) < g(2x)$. Since $2\langle x, y \rangle + 1 > 2x$ for every y , this contradicts the fact that g is a leftmost path. \square

4 Complete separable metric spaces

The following section reviews some of the definitions from [22] that are relevant to the development of the theory of complete separable metric in subsystems of second-order arithmetic. All of the definitions presented here take place in the language of these subsystems.

Definition 4.1 *A (code for a) complete separable metric space \hat{A} consists of a set A together with a pseudometric on A , that is, a function $d : A \times A \rightarrow \mathbb{R}$ such that for all $x, y, z \in A$, $d(x, y) \geq 0$, $d(x, x) = 0$, $d(x, y) = d(y, x)$, and $d(x, z) \leq d(x, y) + d(y, z)$. A (code for a) point of \hat{A} is a sequence $\langle a_n \mid n \in \mathbb{N} \rangle$ of elements of A such that for every n and $m \geq n$ we have $d(a_n, a_m) < 2^{-n}$.*

In other words, a complete separable metric space is presented as the completion of a countable dense subset A , and the elements of such a space are Cauchy sequences with an explicit rate of convergence. Equality and comparisons between reals are defined in a natural way; in particular, the relation $x < y$ is given by a Σ_1 formula, and the relations $x = y$ and $x \leq y$ are Π_1 . A space is said to be *compact* if there is a sequence $\langle F_n \mid n \in \mathbb{N} \rangle$ of finite sequences of points, such that for each n , every point in the space is within a distance of 2^{-n} from some element of F_n .

An *open set* O is presented as the union of a countable sequence of balls, with rational radii, centered at points in A . The notion $x \in O$ is then given by a Σ_1 formula. A *closed set* is presented as the complement of an open set. An

F_σ set is presented as the union of a countable sequence of closed sets, and similarly for G_δ sets, etc. See [22] for more general classes in the Borel and projective hierarchies. A set S is said to be *separably closed* if it is the closure of a countable sequence of elements $\langle y_n \mid n \in \mathbb{N} \rangle$. In other words, a separably closed set S is given by the sequence $\langle y_n \rangle$, and $x \in S$ is defined to mean $\forall \varepsilon > 0 \exists i (d(x, y_i) < \varepsilon)$. Notice that a separably closed subset of a complete separable metric space can be viewed as such a space in its own right, under the inherited metric.

As examples of complete separable metric spaces, \mathbb{R} arises in the usual way as the completion of \mathbb{Q} , and its subspace $[0, 1] \subseteq \mathbb{R}$ is compact. Another example that will be of interest to us is *Baire space*, that is, the space of functions $f : \mathbb{N} \rightarrow \mathbb{N}$, where for $f \neq g$ we define $d(f, g)$ to be $1/2^i$, where i is the least value at which f and g differ. A basis for the topology is given by sets of the form $[\sigma] = \{f \mid f \supseteq \sigma\}$, where σ is any finite sequence of natural numbers. It is not hard to show that Baire space can be represented as the completion of the set of finite sequences of natural numbers with an appropriate metric, and that a closed set in this representation corresponds to the set of paths through a tree. Details can be found in [22].

The definition of a continuous function f between complete separable metric spaces can also be found in [22]; roughly, a continuous function f is given by a sequence of pieces of information to the effect that a ball $B(a, \delta)$ in the domain is mapped into a ball $B(b, \varepsilon)$ in the range. The statement $f(x) = y$ is a Δ_1 statement, and in RCA_0 one can show that if f is a continuous function from one space \hat{A} to another \hat{B} and x is an element of \hat{A} , then there is a unique element y (up to equality in \hat{B}) such that $f(x) = y$. One can also show that the inverse image of an open set is open, and so on.

If S is a (closed, open, etc.) nonempty subset of \hat{A} , $x \in \hat{A}$, and $r \in \mathbb{R}$, then the statement that the distance from x to S is less than r , written $d(x, S) < r$, has to be interpreted as the statement that there is an $\varepsilon > 0$ such that for every $y \in S$, $d(x, y) < r - \varepsilon$. Statements of the form $d(x, S) > r$, $d(x, S) \leq r$, $d(x, S) \geq r$, and $d(x, S) = r$ have to be interpreted in similar ways. In particular, one cannot always show in restricted subsystems that such distances exist, i.e. one cannot always prove $\exists r (d(x, S) = r)$. Note that these notions make sense for any class of elements S that can be defined by a formula in the language of second-order arithmetic.

For S as above, a *locating function* for S is a continuous function f such that for every x , $f(x) = d(x, S)$. In constructive mathematics, it is common to require that sets under consideration are located. We will see in the next section that such functions cannot always be obtained constructively.

We will find it convenient to consider the empty set to be both located and separably closed. In proofs below we can freely proceed by cases according to whether a set in question is empty or not; note that this move is not available in constructive mathematics, where instead one typically restricts one’s attention to sets that are inhabited.

5 Closed sets

The previous section indicates that there are at least four notions of a closed subset of a complete separable metric space that we may reasonably consider:

- (1) closed sets;
- (2) separably closed sets;
- (3) closed and located sets;
- (4) separably closed and located sets.

Furthermore, the metric space in question may or may not be compact. The goal of this section is to clarify the relationships between these four notions, in the context of both compact and arbitrary spaces.

To start with, the following theorem shows that the third and fourth notions are actually equivalent. (This is stated and proved in [8] only for the special case where the space is compact.)

Theorem 5.1 (RCA_0) *Every closed and located set is separably closed and located, and vice-versa.*

Keep in mind that, in the language of second-order arithmetic, there is no notion of an “arbitrary” subset of a metric space. So the theorem above has to be interpreted as the statement that given a (code for) a closed set C and a (code for) a locating function for C , there is a (code for) a separably closed set C' , such that C and C' have the same elements. In other words, the statement asserts the existence of equivalent *representations*. This point should be kept in mind throughout this paper, since we will continue to use such terminological shortcuts.

Proof. If C is separably closed, located, and nonempty, then the set

$$S = \{\langle a, \delta \rangle \in A \times \mathbb{Q} \mid d(a, C) > \delta\}$$

exists by recursive comprehension, and then the complement of C is equal to $\bigcup_{\langle a, \delta \rangle \in S} B(a, \delta)$. This shows that C is closed.

Conversely, suppose C is closed, located, and nonempty. To show that C is separably closed, it suffices to define a countable sequence $\langle x_i \mid i \in \mathbb{N} \rangle$ of elements of C such that for every $a \in A$ and positive $\delta \in \mathbb{Q}$, if $B(a, \delta)$ intersects C then there is an element x_i in $B(a, \delta)$. (Proof: then for every y in C and $\varepsilon > 0$, there is an $a \in A$ such that $d(a, y) < \varepsilon/2$, and an x_i such that $d(x_i, a) < \varepsilon/2$, so that $d(x_i, y) < \varepsilon$.)

Note that there are only countably many such pairs $\langle a, \delta \rangle$, and $B(a, \delta)$ intersects C if and only if $d(a, C) < \delta$. Furthermore, if $B(a, \delta)$ intersects C then there are arbitrarily small values δ' and elements a' of A such that $B(a', \delta') \subseteq B(a, \delta)$ and $B(a', \delta')$ intersects C . (Proof: if y is in $B(a, \delta) \cap C$, then $d(a, y) < \delta$. Given $\delta' < (\delta - d(a, y))/2$, there is an a' in A such that $d(a', y) < \delta'$. So $B(a', \delta')$ intersects C , and $B(a', \delta') \subseteq B(a, \delta)$ since for any $z \in B(a', \delta')$, $d(z, a) \leq d(z, a') + d(a', y) + d(y, a) < \delta' + \delta' + (\delta - 2\delta') = \delta$.)

Now suppose i codes a pair $\langle a, \delta \rangle$ such that $B(a, \delta)$ intersects C . Define a sequence of pairs $\langle b_{i,j}, \gamma_{i,j} \rangle$ by $\langle b_{i,0}, \gamma_{i,0} \rangle = \langle a, \delta \rangle$ and $\langle b_{i,j+1}, \gamma_{i,j+1} \rangle = \langle \hat{b}, \hat{\gamma} \rangle$, where $\langle \hat{b}, \hat{\gamma} \rangle$ is the pair such that $B(\hat{b}, \hat{\gamma}) \subseteq B(b_{i,j}, \gamma_{i,j})$, $B(\hat{b}, \hat{\gamma})$ intersects C , $\hat{\gamma} < 2^{-(j+1)}$, and $\langle \hat{b}, \hat{\gamma} \rangle$ has the least code of any pair satisfying these conditions. Since this sequence can be defined uniformly in i , we can define x_i to be $\langle b_{i,j} \mid j \in \mathbb{N} \rangle$ whenever i codes a pair $\langle a, \delta \rangle$ such that $B(a, \delta)$ intersects C , and any fixed element of C otherwise. Then for every i , $\langle b_{i,j} \mid j \in \mathbb{N} \rangle$ is a Cauchy sequence of elements of A that converges to a point in C , and whenever i codes a pair $\langle a, \delta \rangle$ that intersects C , $\langle b_{i,j} \mid j \in \mathbb{N} \rangle$ converges to a point in C . \square

The following theorem provides an equivalent characterization of closed, located sets (and therefore of separably closed located sets). Roughly speaking, it implies that O is the complement of a closed and located set if and only if the relation $B(a, \delta) \subseteq O$ is decidable, for $a \in A$ and $\delta \in \mathbb{Q}$.

Theorem 5.2 (RCA₀) *A set C is closed and located if and only if the set $S = \{\langle a, \delta \rangle \mid B(a, \delta) \subseteq \overline{C}\}$ exists.*

Proof. The forward direction is included in the proof of the previous theorem. For the converse direction, suppose x is any fixed element of C and suppose that S exists. Since a code for the locating function can be obtained, in RCA₀, from the sequence $\langle d(a, C) \mid a \in A \rangle$, it suffices to show how to define $d(a, C)$ uniformly for $a \in A$. Note that if $\delta = d(a, x) + 1$, then $B(a, \delta)$ intersects C , and so $\langle a, \delta \rangle$ is not in S . For each a we can therefore define $(d_a)_j$ to be the least value of the form $i/2^{j+1}$ such that $B(a, (d_a)_j)$ intersects C , i.e. $\langle a, i/2^{j+1} \rangle \notin S$. Then it is not hard to show that for each a , $\langle (d_a)_j \mid j \in \mathbb{N} \rangle$ is a Cauchy sequence converging to $d(a, C)$. \square

The relationships between closed sets and separably closed sets are expressed by the following theorem are due to Brown [3]. Brown's proof of the reversal of statement 4 to (ACA) is incorrect, but Hirst [11] provides a corrected proof.

Theorem 5.3 (RCA_0) *Each of the following statements is equivalent to (ACA) :*

- (1) *In a compact space, every closed set is separably closed.*
- (2) *In $[0, 1]$, every closed set is separably closed.*
- (3) *In an arbitrary 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 (Π_1^1-CA) :

- (1) *In an arbitrary space, every closed set is separably closed.*
- (2) *In Baire space, every closed set is separably closed.*

What does it take to show that either a closed set or a separably closed set has a locating function? The following theorem answers this question more generally, for classes in the Borel and projective hierarchies as well.

Theorem 5.4 (RCA_0) *Each of the following statements is equivalent to (ACA) :*

- (1) *Every F_σ set in a compact space is located.*
- (2) *Every closed set in a compact space is located.*
- (3) *Every closed set in $[0, 1]$ is located.*
- (4) *Every separably closed set in an arbitrary space is located.*
- (5) *Every separably closed set in $[0, 1]$ is located.*
- (6) *Every open set in an arbitrary space is located.*
- (7) *Every open set in $[0, 1]$ is located.*

Each of the following is equivalent to (Π_1^1-CA) :

- (1) *Every analytic set in an arbitrary space is located.*
- (2) *Every closed set in Baire space is located.*
- (3) *Every G_δ subset of $[0, 1]$ is located.*

Proof. In the first set of statements, the equivalence of 2, 3, and 5 with (ACA) are found in [8], and the equivalence of 4 with (ACA) is found in [7]. Of course, 1 implies 2 and 3, and 4 implies 5. The reversal from 5 to (ACA) also follows from part 6 of Theorem 6.1.

Thus, to show 1–5 are each equivalent to arithmetic comprehension, we only need to prove 1 from (ACA) . To make our argument entirely self-contained, however, we first show that (ACA) implies 4. Suppose C is nonempty and the separable closure of a countable sequence $\langle c_i \mid i \in \mathbb{N} \rangle$. As in the proof of Theorem 5.2 it suffices to show that one can define $d(a, C)$ uniformly for

elements a of the countable dense subset A of the entire space. But using (ACA) this is easy, since $d(a, C) = \inf_i d(a, c_i)$.

To prove 1, we have by Theorem 7.2 below that (ACA) proves that every F_σ subset of a compact space has a separable closure. The desired result follows from this and the fact that (ACA) implies 4, since it is not hard to show that the distance from a point to a set is the same as the distance from the point to its separable closure, assuming the latter exists.

Thus 1–5 are all equivalent to (ACA) . That (ACA) implies 6 follows similarly from the fact that, by Theorem 7.2, (RCA_θ) proves that every open set has a separable closure. Clearly 6 implies 7, and that 7 implies (ACA) follows from part 4 of Theorem 6.1 below.

For the second group of equivalences, the fact that (Π_1^I-CA) implies 1 follows again from the fact that, by Theorem 7.2, (Π_1^I-CA) proves that every analytic set has a separable closure. Clearly 1 implies 2 and 3. The implications from each of 2 and 3 to (Π_1^I-CA) again follow from the stronger results in Theorem 6.1. \square

The equivalence of (Π_1^I-CA) with the statement that every closed set in an arbitrary space is located can also be found in [7]. The more detailed analysis in [8], which focuses on compact spaces, offers additional information about locating functions. For example, if a subset of a compact metric space is both closed and separably closed, a locating function can be obtained in WKL_0 .

The reader may be put off by the forward references to Theorems 7.2 and 6.1 in the proof of Theorem 5.4, but we found it preferable to state all the latter results up front. In the proof, only parts of Theorem 7.2 asserting provability from (ACA) or (Π_1^I-CA) were used, to justify assertions in Theorem 5.4 of the same type; and reverse implications in Theorem 6.1 were used to obtain corresponding reversals in Theorem 5.4. The careful reader can easily verify that there is no circularity.

6 Distances from a point to a set

In this section, we show that the axioms shown by Theorem 5.4 to be necessary for obtaining locating functions are even necessary, in general, to obtain the distance of a single point from a given set.

Theorem 6.1 (RCA_0) *Each of the following statements 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 open subset of $[0, 1]$, then $d(0, O)$ exists.
- (5) In an arbitrary space, if C is a nonempty separably closed set and x is any point, then $d(x, C)$ exists.
- (6) If C is any nonempty separably closed subset of $[0, 1]$, then $d(0, C)$ exists.

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

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

Proof. For the first set of equivalences, provability of 1, 3, and 5 from (ACA) follows easily from parts 2, 6, and 4 of Theorem 5.4, and it is clear that 1, 3, and 5, in turn, imply 2, 4, and 6 of the current theorem, respectively. To show that 2 implies (ACA) , we will use Lemma 2.1. Let a_i be an increasing sequence in $[0, 1]$, and let C be the closed set $\bigcap [a_i, 1]$. Then $\lim a_i = d(0, C)$. To show that 4 implies (ACA) , we can similarly consider the distance from 1 to $\bigcup (0, a_i)$. To show that 6 implies (ACA) , it is not hard to show that the closure of $\bigcup (0, a_i)$ is separably closed; see the proof of Theorem 7.2.1.

For the second cycle of equivalences, provability of 1 and 2 from $(\Pi_1^1\text{-CA})$ again follows from the stronger statements in Theorem 5.4, and clearly 2 implies 3. We are therefore reduced to showing that each of the statements 1 and 3 imply $(\Pi_1^1\text{-CA})$.

For both reversals we will make use of Lemma 3.3. It is well-known that Baire space can be embedded homeomorphically as a G_δ subset of any uncountable Polish space (see [13]). We will show that in fact Baire space, with lexicographic ordering, can be embedded as a G_δ subset of $[0, 1]$ in an order-preserving way. The leftmost path of a non-well-founded tree on ω will then be obtainable from the distance from 0 to a nonempty G_δ subset of $[0, 1]$; and also from the distance from $\lambda x 0$ to a closed subset of Baire space, under the metric induced by the embedding.

The details are as follows. To start with, we need any two increasing sequences $\langle a_i \mid i \in \mathbb{N} \rangle$, $\langle b_i \mid i \in \mathbb{N} \rangle$ in $[0, 1]$ such that

$$0 < a_0 < b_0 < a_1 < b_1 < \dots < 1.$$

To each finite sequence of natural numbers σ assign an open interval A_σ , as follows: $A_\emptyset = (0, 1)$, and once $A_\sigma = (c_\sigma, d_\sigma)$ has been assigned, set $A_{\sigma \cdot \langle i \rangle} =$

$(c_{\sigma^{\langle i \rangle}}, d_{\sigma^{\langle i \rangle}})$, where

$$\begin{aligned} c_{\sigma^{\langle i \rangle}} &= c_\sigma + (d_\sigma - c_\sigma) \cdot a_i \\ d_{\sigma^{\langle i \rangle}} &= c_\sigma + (d_\sigma - c_\sigma) \cdot b_i \end{aligned}$$

Then for each σ and $i < j$, $A_{\sigma^{\langle i \rangle}}$ is to the left of $A_{\sigma^{\langle j \rangle}}$, and if τ is a sequence properly extending σ , A_τ is a subinterval of A_σ with endpoints distinct from those of A_σ . For each i , $b_i - a_i < 1 - a_0 < 1$, and so for each σ , $d_\sigma - c_\sigma \leq (1 - a_0)^{\text{length}(\sigma)}$. This implies that for each element f of Baire space the sequence

$$c_\emptyset, c_{\langle f(0) \rangle}, c_{\langle f(0), f(1) \rangle}, \dots$$

is Cauchy with an explicit rate of convergence, and has the same limit as

$$d_\emptyset, d_{\langle f(0) \rangle}, d_{\langle f(0), f(1) \rangle}, \dots$$

We can associate to each such f the limit of this sequence. The association is clearly injective. In fact, if we define

$$U = \bigcap_{i \in \mathbb{N}} \bigcup_{\{\sigma \mid \text{length}(\sigma) = i\}} A_\sigma,$$

then U is a G_δ set, and each f in Baire space is associated to the limit of

$$\frac{c_\emptyset + d_\emptyset}{2}, \frac{c_{\langle f(0) \rangle} + d_{\langle f(0) \rangle}}{2}, \frac{c_{\langle f(0), f(1) \rangle} + d_{\langle f(0), f(1) \rangle}}{2}, \dots,$$

an element of U . It is not hard to check that conversely, every element of U corresponds to such an f .

Now suppose we are given a tree T on Baire space with at least one path. Then

$$S = \bigcap_{i \in \mathbb{N}} \bigcup_{\{\sigma \in T \mid \text{length}(\sigma) = i\}} A_\sigma$$

is a G_δ set, and the path through T gives rise to an element of S . Assuming 3, the distance d from S to 0 exists. For each i , choose σ_i of length i rightmost in the lexicographic ordering such that $c_{\sigma_i} \leq d$. (There is always such a σ_i , since there is at least one path through T ; if g is such a path, τ is the initial segment of g of length i , and τ' is to the right of τ , then $c_{\tau'} > d$.) Then for each i , there is no point in S to the left of A_{σ_i} , since $d(0, S) \geq c_{\sigma_i} > c_\tau$, for any τ to the left of σ_i . On the other hand, for each i there is a point of S in A_{σ_i} , since otherwise $d < c_\tau$ for the τ of length i immediately to the right of σ_i . In particular, for each $i < j$ we have that σ_j extends σ_i , and the element of Baire space corresponding to $\lim c_{\sigma_i}$ is the leftmost path through T . This shows that 3 implies $(\Pi_1^1\text{-CA})$.

To show that 1 implies $(\Pi_1^1\text{-CA})$, note that in the construction above, Baire space (corresponding to U) can be viewed as a complete separable metric space

in its own right with the induced metric from $[0, 1]$. In this subspace the collection of paths through a tree T is a closed set S , and the leftmost path through a non-well-founded tree is the path closest to the constant zero sequence $\lambda x 0$. As above, this leftmost path can be computed from $d(\lambda x 0, S)$. \square

7 Closures

The *closure* of a set X is defined to be a closed set that is included in every closed set that includes X , and the *separable closure* of a set X is defined to be a separably closed set that is included in every separably closed set that includes X . (The *interior* of X is defined similarly.) Given the information we have about closed and separably closed sets, a reasonable question to ask is: what does it take to obtain the closure, or separable closure, of a set in the Borel hierarchy?

Theorem 7.1 (RCA₀) *Each of the following statements is equivalent to (ACA):*

- (1) *In a compact space, every F_σ set has a closure (or, dually, every G_δ has an interior).*
- (2) *In an arbitrary space, every open set has a closure (or, every closed set has an interior).*
- (3) *In $[0, 1]$, every open set has a closure (or, every closed set has an interior).*

Each of the following statements is equivalent to (Π_1^1 -CA):

- (1) *In an arbitrary space, every analytic set has a closure (or every co-analytic set has an interior).*
- (2) *In Baire space, every F_σ set has a closure (or every G_δ set has an interior).*
- (3) *In $[0, 1]$, every G_δ set has a closure (or every G_δ set has an interior).*

Proof. Let us consider the first set of equivalences first. To show that (ACA) implies 1, suppose that $S = \bigcup C_i$ in a compact space $X = \hat{A}$, with each C_i closed. By Theorem 5.3 each C_i is separably closed, and so includes a countable dense sequence $\langle x_{i,j} \mid j \in \mathbb{N} \rangle$. But then if $B = B(a, \delta)$ is the open ball around a with radius δ , with a in A and δ rational, B is a subset of the complement of S if and only if for every i and j , $d(a, x_{i,j}) \geq \delta$. This last condition is arithmetic, and the union of these $B(a, \delta)$ is the interior of the complement of S . So, using arithmetic comprehension, the closure of S exists.

The proof that (ACA) implies 2 is similar; if $O = \bigcup B(b_i, \varepsilon_i)$ is open, $B(a, \delta)$ is a subset of the complement of O if and only if for every i , $d(a, b_i) > \delta + \varepsilon_i$.

Clearly each of 1 and 2 imply 3. We only need to show that 3 implies arithmetic comprehension. To that end, let a_i be an increasing sequence as in Lemma 2.1, let $O = \bigcup(0, a_i)$, and let C be the closure of O . If there is a rational number q such that q is the limit of the a_i we are done. Otherwise, the collection of rationals in O is the same as the collection of nonzero rationals in C ; since this provides both a Σ_1 and Π_1 description, this collection forms a set. It is easy to obtain the limit of the a_i computably from this set, that is, using recursive comprehension.

For the second cycle of equivalences, the proof that $(\Pi_1^1\text{-CA})$ proves that every analytic set has a closure is similar to the proofs of 1 and 2 above. If S is analytic, then $B(a, \delta)$ is a subset of the complement of S if and only if

$$\forall y (d(a, y) < \delta \rightarrow y \notin X),$$

and this last formula is Π_1^1 . So the set of $a \in A$ and $\delta \in \mathbb{Q}$ with this property, and hence the interior of the complement of S , can be obtained using Π_1^1 comprehension.

Clearly 1 implies 2 and 3. To show that 2 implies Π_1^1 comprehension, first note that we have arithmetic comprehension by the first part of the theorem. We will use Lemma 3.2. Let T_i be any sequence of trees on ω ; the idea is that by constructing an appropriate F_σ set we can use the closure to test whether or not there is a path through each T_i .

Specifically, for each i and j , let $S_{i,j}$ be the tree

$$S_{i,j} = \{\sigma \mid \sigma \subseteq \langle i, 0, \dots, 0 \rangle\} \cup \{\langle i, 0, \dots, 0 \rangle \hat{\ } \sigma \mid \sigma \in T_i\}$$

where there are j zeros after i in the indicated sequence. In other words, $S_{i,j}$ is a tree obtained by grafting a copy of T_i onto an initial segment $\langle i, 0, \dots, 0 \rangle$.

Let $C_{i,j}$ be the closed set of paths through $S_{i,j}$. Then for every j , $C_{i,j}$ is nonempty if and only if there is a path through T_i , and so the infinite sequence $\langle i, 0, 0, \dots \rangle$ is in the closure of $\bigcup_j C_{i,j}$ if and only if there is a path through T_i . Let $D = \bigcup_{i,j} C_{i,j}$ and, assuming 2, let E be the closure of D . Then for each i , there is a path through T_i if and only if $\langle i, 0, 0, 0, \dots \rangle$ is in E . Hence the set of i such that there is a path through T_i exists, by arithmetic comprehension.

Finally, to show that 3 implies Π_1^1 comprehension, we will show that the embedding of Section 6 sends the set E just constructed to a G_δ subset of $[0, 1]$. Let S be the union of the trees $S_{i,j}$. The set of paths through S is a closed subset of Baire space; as in Section 6 this embeds as a G_δ subset of $[0, 1]$. The set $\bigcup_{i,j} C_{i,j}$ is equal to the set of paths through S minus a countable set of paths of the form $\langle i, 0, 0, \dots \rangle$. The latter embeds as a countable set of points, which is therefore F_σ . Thus the embedding of $\bigcup_{i,j} C_{i,j}$ is equal to a G_δ set minus an F_σ set, and hence G_δ . As above, there is a path through T_i if and only

if the embedding of $\langle i, 0, 0, \dots \rangle$ is in the closure of the embedding of $\bigcup_{i,j} C_{i,j}$. So, by (ACA), if the latter closure exists, then so does the set of i such that there is a path through T_i . \square

Theorem 7.2 (RCA₀) *Every open subset of an arbitrary space has a separable closure. Also, each of the following statements is equivalent to (ACA):*

- (1) *Every F_σ subset of a compact space has a separable closure.*
- (2) *Every closed subset of $[0, 1]$ has a separable closure.*

Each of following statements is equivalent to (Π_1^1 -CA):

- (1) *Every analytic subset of an arbitrary space has a separable closure.*
- (2) *Every closed subset of Baire space has separable closure.*
- (3) *Every G_δ subset of $[0, 1]$ has a separable closure.*

Proof. For the first statement, it is easy to see that if $B(a, \delta)$ is any open ball and A is the countable dense subset of the entire space, then $B(a, \delta) \cap A$ is a countable dense subset of $B(a, \delta)$. So in RCA₀ the closure of any open ball is the closure of a sequence of points from A . It is also not hard to show that the closure of a union of open balls is the closure of the union of the corresponding dense subsets. The argument that ACA₀ proves that every F_σ subset of a compact space has a separable closure is similar, given that, by Theorem 5.3, ACA₀ proves that every closed subset of a compact space is separably closed. The reversal from 2 to (ACA) is also given by Theorem 5.3, since if a closed set C has a separable closure C' , it is not hard to see that $C = C'$.

In the next set of equivalences, that (Π_1^1 -CA) implies 1 follows from the fact that (Π_1^1 -CA) proves that every analytic set has a closure (by Theorem 7.1) and that every every closed set is separably closed (by Theorem 5.3). Clearly 1 implies 2 and 3. That 2 implies (Π_1^1 -CA) follows from Theorem 5.3.

For the last reversal, suppose 3. By the first part of the theorem, we have (ACA), and hence, by Theorem 5.4, that every set with a separable closure is located. Thus 3 implies that every G_δ subset of $[0, 1]$ is located, which, again by Theorem 5.4, implies Π_1^1 -CA. \square

8 Iterative functions on metric spaces

In numerical and functional analysis, it is common to define functions, or sequences of elements of a space, using iterative procedures that can be given a computational interpretation. For example, if $f(x)$ is a computable (and hence continuous) function from \mathbb{R} to \mathbb{R} , then the function $g(n, x)$ returning the n th

iterate $f^n(x)$ of f on x is computable. It follows that for any x the sequence $\langle x_n \mid n \in \mathbb{N} \rangle$ defined by $x_n = g(n, x)$ is also computable. One has to be careful in subsystems of second-order arithmetic, however, since verifying that the resulting function or sequence is well-defined can require some axiomatic strength. For example, Friedman, Simpson, and Yu [6] show that either Σ_2 induction or (*WKL*) can be used to show that if f is continuous on a compact space, then so is every iterate $f^n(x)$; and that this latter statement is in fact *equivalent* to the disjunction of these two principles.

Friedman, Simpson, and Yu note, however, that the statement is provable in RCA_0 in cases where f has a *modulus of uniform continuity*, i.e. there is a function $g : \mathbb{N} \rightarrow \mathbb{N}$ such that for every $n \in \mathbb{N}$, x , and y , if $d(x, y) < 2^{-g(n)}$, then $d(f(x), f(y)) < 2^{-n}$. The next theorem is a useful generalization of this fact, justifying a strong form of primitive recursion.

We view functions $f(x_0, \dots, x_{k-1})$ with multiple arguments as functions on the associated product space, so saying that g is a modulus of continuity means that for every n , \vec{x} , and \vec{y} , if $d(x_i, y_i) < 2^{-g(n)}$ for all $i < k$, then $d(f(\vec{x}), f(\vec{y})) < 2^{-n}$. In the statement of the theorem, \vec{z} is meant to be a sequence of parameters from any choice of separable metric spaces, which can include e.g. \mathbb{N} as a discrete space.

Theorem 8.1 (RCA_0) *Let $f_0(\vec{z}), f_1(\vec{z}, x_0), f_2(\vec{z}, x_0, x_1), \dots$ be any sequence of continuous functions with arguments from complete separable metric spaces, where the functions f_i have a common range, X , and the arguments x_0, x_1, \dots are from X . Assume also that there is a sequence g_0, g_1, g_2, \dots such that each g_i is a modulus of uniform continuity for f_i . Then there is a sequence of continuous functions $h_i(\vec{z})$ satisfying*

$$\begin{aligned} h_0(\vec{z}) &= f_0(\vec{z}) \\ h_{n+1}(\vec{z}) &= f_{n+1}(\vec{z}, h_0(\vec{z}), \dots, h_n(\vec{z})). \end{aligned}$$

Proof. We simply modify the proof of [6, Lemma 2.7], adopting the notation $(a, r)f(b, s)$ used there to mean that the code for the continuous function f maps $B(a, r)$ into $B(b, s)$. Keep in mind that here the relevant metrics are sup metrics on product spaces.

As in [6] we can assume without loss of generality that all the codes f_k satisfy the property that for all $k, n \in \mathbb{N}$ and $B(\vec{a}, r)$ with $r < 2^{-g_k(n)}$ there is a neighborhood condition $(\vec{a}, r)f_k(b, s)$ with $s < 2^{-n}$. Define the functions h_n so that $(\vec{a}, r)h_n(\vec{b}, s)$ holds if and only if there are sequences b_0, \dots, b_{n-1} in the common range of the functions f_i and s_0, \dots, s_{n-1} in \mathbb{Q} such that $r \leq s_0 \leq \dots \leq s_{n-1}$ and

$$(\vec{a}, r)f_0(b_0, s_0), (\vec{a} \hat{\ } \langle b_0 \rangle, s_0)f_1(b_1, s_1), \dots, (\vec{a} \hat{\ } \langle b_0, \dots, b_{n-1} \rangle, s_{n-1})f_n(b, s).$$

As in [6], using the moduli of uniform continuity for the functions f_n it is not hard to show that each h_n is a code for a continuous function that is everywhere defined, and that the sequence $\langle h_i \rangle$ satisfy the defining equations. \square

In fact, one can obtain moduli of uniform continuity for the sequence of functions h_n , though we will not need this below. Note that if there are no parameters \vec{z} , the result of the lemma is just a sequence h_0, h_1, h_2, \dots of elements of the underlying space.

9 Hilbert spaces and Banach spaces

We now turn to the general theory of Banach spaces, and the theory of Hilbert spaces as a particular case. In light of Sections 5–7 and the fact that infinite dimensional Hilbert and Banach spaces are not even locally compact, dealing with distances, closures, and countable dense subsets may, a priori, require strong axioms. Our goal will be to understand how, in some cases, we can use the additional structure of a Hilbert or Banach space to avoid using the full strength of $(\Pi_1^1\text{-CA})$. The definitions that follow are from [22].

Definition 9.1 *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 9.2 *A (real) separable Banach space B consists of a countable vector space A over \mathbb{Q} together with function $\|\cdot\| : A \rightarrow \mathbb{R}$ satisfying*

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

Given a Banach space B as above, we can define a pseudometric $d(x, y)$ on A by $d(x, y) = \|x - y\|$. We think of B as the completion of A under this pseudometric, and often write $B = \hat{A}$. We thus define an *element* x of B to be a sequence $\langle x_n \mid n \in \mathbb{N} \rangle$ of elements of A , such that $d(x_n, x_m) < 2^{-n}$ whenever $n < m$. The norm is extended to the whole space by defining $\|x\| = \lim_n \|x_n\|$ for $x = \langle x_n \rangle$. We define $x = y$ to mean that $d(x, y) = 0$, making d a metric. As a metric space, then, B is separable and Cauchy complete, provably in RCA_0 .

Definition 9.3 *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 $a, b \in \mathbb{Q}$.

Every Hilbert space can be viewed as a Banach space with norm $\|x\| = \langle x, x \rangle^{1/2}$. The triangle inequality and the Cauchy-Schwartz inequality follow from the axiomatic characterization of the inner product, and then the inequality

$$\|\langle x, y \rangle - \langle z, w \rangle\| = \|\langle x, y - w \rangle + \langle x - z, w \rangle\| \leq \|x\| \|y - w\| + \|x - z\| \|w\|$$

shows that the inner product is continuous. We view $H = \hat{A}$ as the completion of A as above, and extend the inner product to the whole space by defining $\langle x, y \rangle = \lim_n \langle x_n, y_n \rangle$ for $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$; the inequality above can be used to find an explicit code for the inner product as a continuous function on $H \times H$.

The standard examples of separable infinite dimensional Hilbert spaces as well as of Banach spaces can be developed in 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 (see, for example, [22,4]). Examples of Banach spaces are the space $C(X)$ of continuous functions on a compact space X under the sup norm, and $L^1(X)$, the space of integrable real-valued functions over X (see [22]).

In the theory of Banach spaces, an important notion is that of a bounded linear operator:

Definition 9.4 *A bounded linear operator between separable Banach spaces \hat{A} and \hat{B} , is a function $F : A \rightarrow \hat{B}$ such that*

- (1) *F is linear, i.e. $F(q_1 a_1 + q_2 a_2) = q_1 F(a_1) + q_2 F(a_2)$ for all $q_1, q_2 \in \mathbb{Q}$ and $a_1, a_2 \in A$.*
- (2) *The norm of F is bounded, i.e. there exists a real number M such that $\|F(a)\| \leq M \|a\|$ for all $a \in A$.*

Then, for $x = \langle x_n \mid n \in \mathbb{N} \rangle \in \hat{A}$ we define $F(x) = \lim_{n \rightarrow \infty} F(x_n)$.

The fact that the limit defining $F(x)$ exists can be established in RCA_0 . Note that if M satisfies the second clause in the definition then $\|F(x)\| \leq M \|x\|$ for every $x \in \hat{A}$. The infimum of all such M , if it exists, is called the *norm* of F , and denoted by $\|f\|$.

Simpson [22] shows that every bounded linear operator on a Banach space is a continuous function on the associated metric space, and, conversely, every continuous and linear function is a bounded linear operator. We will, in particular, be interested in bounded linear *functionals*, that is, bounded linear operators from a Banach space to \mathbb{R} .

A number of good introductory textbooks cover the elementary facts about Hilbert and Banach spaces that we will use here; [5,17] are two examples.

10 Bases and independent generating sequences

With respect to a Banach space B , we will call a sequence $S = \langle x_i \mid i \in \mathbb{N} \rangle$ an *independent generating sequence* if its elements are linearly independent (that is, zero cannot be written nontrivially as a finite linear combination of elements from S) and B is the closure of the linear span of S . Note that when B is the closure of the linear span of S , the sequence of all rational linear combinations of elements of S is dense in B .

In this section we will show that WKL_0 proves that every Banach space has an independent generating sequence. The corresponding statement for Hilbert space can be obtained in RCA_0 . In fact, RCA_0 proves that there is a generating sequence in which the elements are *orthonormal*, which is to say, distinct elements are orthogonal (i.e. have inner product 0) and the norm of each element is 1. The usual proof shows, in RCA_0 , that such a sequence is necessarily a (*Schauder*) *basis*, which is to say, every element of the space can be represented uniquely as an infinite linear combination of basis elements. (It is not the case that every Banach space has a Schauder basis.)

Our constructions will use a result from the development of Banach spaces in [12], and draw on ideas from the corresponding developments in recursive mathematics [20] and constructive mathematics [1]. There are, however, subtleties and key differences, some of which are discussed at the end of this section.

Given a Banach space $B = \hat{A}$, by applying the law of the excluded middle in RCA_0 we have that either some finite sequence of elements of A spans all of A , or not. In the first case, B is said to be *finite dimensional*, and in the second case, B is said to be *infinite dimensional*. These definitions will be further justified by developments below. Note that this appeal to the law of the excluded middle is not available in constructive mathematics, a fact which accounts for many of the differences between the presentations of [1] and [20].

First we will show that RCA_0 proves that every finite dimensional Hilbert space has an orthonormal basis, and that WKL_0 proves that every finite dimensional Banach space has an independent generating sequence. Then we will consider the infinite cases.

Lemma 10.1 *Assuming that $\langle v_0, \dots, v_k \rangle$ is a sequence of vectors in a Hilbert space, the statement that the given vectors are linearly independent is equiva-*

lent to a Σ_1 formula in RCA_0 .

Proof. Imagine using Gram-Schmidt process on v_0, \dots, v_k to obtain an orthonormal sequence e_0, \dots, e_k : define

$$e_0 = v_0 / \|v_0\|$$

$$e_{i+1} = \frac{v_{i+1} - \sum_{j \leq i} \langle v_{i+1}, e_j \rangle e_j}{\|v_{i+1} - \sum_{j \leq i} \langle v_{i+1}, e_j \rangle e_j\|}.$$

The process comes to a screeching halt at some stage i if and only if the numerator of the corresponding fraction is 0, which happens, in turn, if and only if v_i is linearly dependent on v_0, \dots, v_{i-1} .

It is the fact that there is a quotient involved that prevents us from applying Theorem 8.1 to define the sequence e_0, \dots, e_k by primitive recursion up to k ; otherwise, functions obtained by composing the inner product, finite sums and products of reals, and so on, have moduli of uniform continuity (see [22, Remark IV.2.8]). The solution is to modify the construction to accept a rational parameter ε and make sure that we never divide by a quotient less than ε , thereby guaranteeing an appropriate modulus of continuity. In other words, we use primitive recursion up to k to define a sequence e'_0, e'_1, \dots, e'_k , depending on ε , satisfying

$$e'_0 = v_0 / \max(\varepsilon, \|v_0\|)$$

$$e'_{i+1} = \frac{v_{i+1} - \sum_{j \leq i} \langle v_{i+1}, e'_j \rangle e'_j}{\max(\varepsilon, \|v_{i+1} - \sum_{j \leq i} \langle v_{i+1}, e'_j \rangle e'_j\|)}.$$

We claim that v_0, \dots, v_k are linearly independent if and only if

for some rational $\varepsilon > 0$, for every $i \leq k$, $\|v_i - \sum_{j < i} \langle v_i, e'_j \rangle e'_j\| > \varepsilon$.

Using Σ_1 collection, derivable in RCA_0 (see [22,9]), this formula is equivalent to one that is Σ_1 .

First, note that, assuming for some $\varepsilon > 0$ the property holds for every i less than or equal to some value $l \leq k$, the definition of the sequence e'_0, \dots, e'_l amounts to the definition of e_0, \dots, e_l . Using induction up to l we can show that the sequence is orthonormal, i.e. $\|e_i\| = 1$ for every $i \leq l$, and $\langle e_i, e_j \rangle = 0$ for every $j < i \leq l$. Using primitive recursion up to l (on the space of finite sequences of reals, i.e. a disjoint union of the spaces \mathbb{R}^i for $i \in \omega$), we can also work backwards to solve for the e_i 's in terms of the v_i 's. In other words, we can compute a sequence d_0, d_1, d_2, \dots such that each d_i is a sequence of reals of length $i + 1$, and such that we can show by induction up to k that for each $i \leq l$ we have

$$e_i = \sum_{j \leq i} (d_i)_j v_j.$$

To show the formula above works as advertised, suppose, in the forward direction, there is an ε with this property, but the sequence v_0, \dots, v_k is linearly dependent. By the preceding paragraph, we have that e_0, \dots, e_k are linearly dependent, so there is a sequence $\alpha_0, \dots, \alpha_k$, at least one of which is nonzero, such that

$$\alpha_0 e_0 + \dots + \alpha_k e_k = 0.$$

If $\alpha_i \neq 0$, we obtain the usual contradiction by taking the inner product of both sides with e_i .

On the other hand, suppose there is no such ε . Let $\psi(m)$ be the formula

$$\exists \varepsilon > 0 \forall i \leq m (\|v_i - \sum_{j < i} \langle v_i, e'_j \rangle e'_j\| > \varepsilon).$$

We are assuming $\neg\psi(k)$; by the least element principle, we obtain the least value of $m \leq k$ such that $\neg\psi(m)$. So we have

$$(\forall j < m \psi(j)) \wedge \neg\psi(m).$$

Using Σ_1 collection on the left conjunct, we obtain a single ε that works uniformly; hence, by the discussion there is a sequence of coefficients expressing v_0, \dots, v_{m-1} in terms of e_0, \dots, e_{m-1} . On the other hand, we have that for every $\varepsilon > 0$,

$$\|v_m - \sum_{j < m} \langle v_m, e'_j \rangle e'_j\| < \varepsilon.$$

so $v_m - \sum_{j < m} \langle v_m, e'_j \rangle e'_j$ is equal to 0. Thus there is an expression for v_m in terms of v_0, \dots, v_{m-1} , showing that v_0, \dots, v_m are not linearly independent. \square

Lemma 10.2 (RCA₀) *Every finite dimensional Hilbert space has an orthonormal basis.*

Proof. Let σ be a finite sequence of elements of A that spans A , and suppose A has length m . Let $\varphi(n)$ be a formula asserting

There is a sequence $\langle v_0, \dots, v_{n-1} \rangle$ of distinct elements of σ that is linearly independent.

Applying the least-element principle to the formula $\theta(i) \equiv \varphi(m-i)$ we obtain a maximum value of k such that there is a linearly independent sequence of elements of σ of length k . Let $S = \langle v_0, v_1, \dots, v_{k-1} \rangle$ be such a sequence. By maximality, S spans σ , and hence A . Since S is independent, there is an $\varepsilon > 0$ such that the Gram-Schmidt process goes through, yielding $\langle e_0, \dots, e_{k-1} \rangle$. By the argument above, one can prove, in RCA₀, that this sequence is orthonormal, and spans the elements $\langle v_0, \dots, v_{k-1} \rangle$. \square

In fact, the argument above can be used to show that any two bases for a finite dimensional space have the same cardinality, and so the notion of dimension

for such spaces is well-defined.

Our proof of Lemma 10.2 made heavy use of Σ_1 induction. This is unavoidable, since [22, page 411] cites unpublished work by Friedman that shows that the statement that every finite dimensional vector space has a basis implies Σ_1 induction over a weak base theory.

More is required to handle more general Banach spaces. The following lemma is a consequence of Humphreys and Simpson [12]; see also the “independence criterion” in [20, page 143], or [17, Lemma 2.4-1].

Lemma 10.3 *Assuming that $\langle v_0, \dots, v_k \rangle$ is a sequence of vectors in a Banach space, the statement that the given vectors are linearly independent is equivalent to a Σ_1 formula in WKL_0 .*

Since independence of the given vectors is easily expressible with a Σ_2 formula, and the Σ_2 least-element principle can be justified from Σ_2 induction, we obtain the following just as in the proof of Lemma 10.2:

Lemma 10.4 (RCA_0) *The disjunction of (WKL) and Σ_2 induction implies that every finite dimensional Banach space has an independent generating sequence.*

We now turn to the infinite dimensional case. The proof below adapts a construction from [20, Section 4.7], carried out there in the context of a Banach space.

Lemma 10.5 (RCA_0) *Every infinite dimensional Hilbert space has an independent generating sequence.*

Proof. Elements of our generating sequence will be chosen from the underlying vector space, A . We start by choosing an enumeration of the natural numbers u_0, u_1, u_2, \dots such that each natural number occurs infinitely often; for example, we can define $u_i = (i)_0$, where $(i)_0$ denotes the first number of the pair coded by i in any reasonable coding of pairs of natural numbers.

Suppose at stage i we have already chosen elements $x_{j_0}, x_{j_1}, \dots, x_{j_k}$ to be part of the generating sequence. We next consider whether to add x_{u_i} . By the law of the excluded middle, either $\langle x_{j_0}, \dots, x_{j_k}, x_{u_i} \rangle$ is linearly independent, or it isn't. In the first case, there is a witness to the Σ_1 formula characterizing linear independence given by Lemma 10.1; in the second case, there is a finite sequence of rationals q_0, \dots, q_k such that

$$\|x_{u_i} - \sum_{l \leq k} q_l x_{j_l}\| < 2^{-i}.$$

Hence, if we search simultaneously for a suitable witness to the independence

and a suitable sequence of rationals (together with a witness to the corresponding inequality), we are guaranteed to find one or the other. If we find the witness to the independence, we add x_{u_i} to the generating sequence, by defining j_{k+1} to be u_i ; otherwise, we do nothing. Notice that in the second case, we have not established that x_{u_i} is linearly dependent on the previous vectors, but only that it can be approximated to within a factor of 2^{-i} . This is why we chose our enumeration $\langle u_i \rangle$ in such a way that each element x_n is considered infinitely often; in the end, if x_n has not been added at any stage, it is because it is in the closure of the span of our generating sequence.

Let us describe the proof in a little more detail. By Lemma 2.2 there is a function $f(\sigma, i)$ with the following property: whenever σ is a finite sequence $\langle \sigma_0, \dots, \sigma_k \rangle$ of indices, $f(\sigma)$ either returns a number witnessing the fact that $\langle x_{\sigma_0}, \dots, x_{\sigma_k} \rangle$ is linearly independent, or a sequence of rationals q_0, \dots, q_{k-1} such that $\|x_{\sigma_k} - \sum_{l \leq k} q_l x_{\sigma_l}\| < 2^{-i}$ together with a witness to this inequality.

Define a function j by primitive recursion on the natural numbers by setting $j(0)$ to be the empty sequence, and setting $j(i+1)$ to be equal to $j(i) \hat{\ } i$ if $f(j(i) \hat{\ } \langle x_{u_i} \rangle, i)$ returns a witness for linear independence, and $j(i)$ otherwise.

Because we are assuming that no finite sequence of elements of A spans all the elements of A , we have that for every i there is an $n > i$ such that $j(n)$ properly extends $j(i)$. Define a function k by primitive recursion, where $k(0)$ is the least value n such that $j(n)$ is not the empty sequence, and $k(i+1)$ is the least value of n such that $j(n)$ properly extends $j(i)$. We obtain the desired generating sequence by defining $v_i = x_{(j(k(i)))_i}$. \square

The argument just given can be generalized to arbitrary Banach spaces, Using the Σ_1 independence criterion given by Lemma 10.3, rather than the Σ_1 criterion for Hilbert spaces. However, Lemma 10.3 relied on (*WKL*). In fact, the argument can be made to go through in RCA_0 , though this involves unraveling some of the technical details of the independence criterion. The proof below relies on the following notation from [20, page 143]: for each m and k , let S_{mk} denote the set of all k -tuples $\langle \beta_0, \dots, \beta_{k-1} \rangle$ of rationals whose denominators are 2^m and which satisfy

$$1 \leq |\beta_0|^2 + \dots + |\beta_{k-1}|^2 \leq 4.$$

Lemma 10.6 (RCA_0) *Suppose x_0, \dots, x_k are elements of a Banach space, and suppose, for some m ,*

$$\min\{\|\beta_0 x_0 + \dots + \beta_{k-1} x_{k-1}\| \mid \langle \beta_0, \dots, \beta_{k-1} \rangle \in S_{mk}\} > 2^{-m}(\|x_0\| + \dots + \|x_{k-1}\|).$$

Then, for any $\varepsilon > 0$, either there is a sequence of rationals q_0, \dots, q_{k-1} such

that

$$\|x_k - \sum_{i < k} q_i x_i\| < \varepsilon,$$

or there is an m' such that

$$\min\{\|\beta_0 x_0 + \dots + \beta_k x_k\| \mid \langle \beta_0, \dots, \beta_k \rangle \in S_{m'(k+1)}\} > 2^{-m'}(\|x_0\| + \dots + \|x_k\|).$$

Furthermore, if the second disjunct holds, the sequence $\langle x_0, \dots, x_{k-1} \rangle$ is independent.

Proof. Given m and ε as in the hypotheses, let us work backwards to determine conditions on a choice of m' that guarantee the conclusion.

First, note that the condition on m guarantees that for any $m' \geq m$ and any $\langle \beta'_0, \dots, \beta'_{k-1} \rangle \in S_{m'k}$, we have

$$\|\beta'_0 x_0 + \dots + \beta'_{k-1} x_{k-1}\| > \delta(\|x_0\| + \dots + \|x_{k-1}\|)$$

with $\delta = (4/9)2^{-(m+1)} > 2^{-(m+3)}$. This is so because given any such sequence we can scale it by a factor of at most $3/2$ to obtain a sequence of real numbers $\langle \gamma_0, \dots, \gamma_{k-1} \rangle$ on the sphere of radius $3/2$, i.e. satisfying $|\gamma_0|^2 + \dots + |\gamma_{k-1}|^2 = 9/4$, and then approximate $\langle \gamma_0, \dots, \gamma_{k-1} \rangle$ by a sequence $\langle \beta_0, \dots, \beta_{k-1} \rangle$ in S_{mk} with each $|\beta_i - \gamma_i| \leq 2^{-m+1}$.

Now, for any m' , if the second disjunct fails, there is a sequence $\langle \beta_0, \dots, \beta_{k-1} \rangle$ in $S_{m'(k+1)}$ such that

$$\|\beta_0 x_0 + \dots + \beta_k x_k\| \leq 2^{-m'}(\|x_0\| + \dots + \|x_k\|). \quad (1)$$

Assuming $\beta_k > 0$ and $2^{-m'}/\beta_k < \varepsilon$, we can divide through by β_k to obtain the desired sequence q_0, \dots, q_{k-1} . It therefore suffices to obtain a lower bound for β_k for sufficiently large m' in terms of m and x_0, \dots, x_{k+1} , since then we can simply choose m' large enough to ensure $2^{-m'}/\beta_k < \varepsilon$.

Assume, then, that (1) holds. Using the triangle inequality in the form $\|u\| \geq \|v\| - \|u + v\|$, we have

$$\begin{aligned} \|\beta_k x_k\| &\geq \|\beta_0 x_0 + \dots + \beta_{k-1} x_{k-1}\| - \|\beta_0 x_0 + \dots + \beta_k x_k\| \\ &> 2^{-(m+3)}(\|x_0\| + \dots + \|x_{k-1}\|) - 2^{-m'}(\|x_0\| + \dots + \|x_k\|). \end{aligned}$$

If $m' \geq m + 4$, we have

$$\begin{aligned} (\beta_k + 2^{-m'})\|x_k\| &> (2^{-(m+3)} - 2^{-m'})(\|x_0\| + \dots + \|x_{k-1}\|) \\ &\geq 2^{-(m+4)}(\|x_0\| + \dots + \|x_{k-1}\|) \end{aligned}$$

and so

$$\begin{aligned}\beta_k &> (2^{-m+4}/\|x_k\|)(\|x_0\| + \dots + \|x_{k-1}\|) - 2^{-m'} \\ &> (2^{-m+5}/\|x_k\|)(\|x_0\| + \dots + \|x_{k-1}\|)\end{aligned}$$

as long as m' is large enough.

The fact that the second disjunct implies that the sequence $\langle x_0, \dots, x_{k-1} \rangle$ is independent is proved as in [20]; the argument is similar to that in the second paragraph in this proof. \square

Since the inequalities in the statement of Lemma 10.6 are Σ_1 , making the corresponding changes to the proof of Lemma 10.5 shows the following:

Lemma 10.7 (RCA₀) *Every infinite dimensional Banach space has an independent generating sequence.*

In the case of Hilbert spaces, we can apply the Gram-Schmidt process to make any generating sequence orthonormal.

Lemma 10.8 (RCA₀) *Every infinite dimensional Hilbert space has an orthonormal basis.*

Proof. Let v_0, v_1, \dots be an independent generating sequence for the space. Since each initial segment v_0, \dots, v_k is independent, as a by-product of Lemma 10.1 we know that for each k there is a ε such that in the Gram-Schmidt process run up to k , the norm in the denominator is greater than ε . This last expression is Σ_1 , so by Lemma 2.2 there is a function $h(i)$ such that for each i , the norm in the denominator is at least $h(i)$. Thus we can re-express the full Gram-Schmidt process with the recursion

$$\begin{aligned}e_0 &= v_0 / \max(h(0), \|v_0\|) \\ e_{i+1} &= \frac{v_{i+1} - \sum_{j \leq i} \langle v_{i+1}, e_j \rangle e_j}{\max(h(i+1), \|v_{i+1} - \sum_{j \leq i} \langle v_{i+1}, e_j \rangle e_j\|)}.\end{aligned}$$

It is the presence of h in the denominator that ensures that each step of the process has a modulus of uniform continuity, thereby allowing us to apply Theorem 8.1. As before we can employ a complementary recursion to re-express the v_i 's in terms of the e_i 's, and use induction to show that the sequence is orthonormal and spans the v_i 's.

The usual proofs can be carried out in RCA₀ to show that if x is any element of the space then $x = \sum_i \langle x, e_i \rangle e_i$, and that this representation is unique. \square

Summing up, then, we have shown:

Theorem 10.9 (RCA₀) *Every Hilbert space has an orthonormal basis.*

Theorem 10.10 RCA_0 together with the disjunction of (WKL) and Σ_2 induction proves that every Banach space has an independent generating sequence.

The nonconstructive components in the proofs of these two theorems include the use of the law of the excluded middle to distinguish the finite-dimensional and infinite-dimensional cases and the use of the least-element principle in the finite-dimensional case. Thus, one obtains constructive versions if one either restricts to the infinite dimensional case, or if one allows vectors in the “orthonormal basis” to be zero. One can also obtain a constructive version of Lemma 10.2 with a more stringent notion of a finite dimensional space. Details can be found in [1].

Theorems 10.9 and 10.10 imply their computable analogs, which is what is addressed by Pour-El and Richards [20]. In other words, the constructions described are computable uniformly from the underlying space (though, of course, the dimension of the space cannot be determined computably, in general); it is just the axiomatic verification that the construction works that requires extra effort.

Two Hilbert spaces H_1, H_2 are said to be *isomorphic* if there exist bounded linear functionals $F : H_1 \rightarrow H_2$ and $G : H_2 \rightarrow H_1$ that are inverses to each other. If $\langle e_i \rangle$ and $\langle e'_i \rangle$ are orthonormal bases of H_1 and H_2 respectively of the same cardinality, clearly we can obtain such an F and G by setting $F(e_i) = e'_i$ and $G(e'_i) = e_i$.

Corollary 10.11 (RCA_0) *Any two infinite dimensional Hilbert spaces are isomorphic, as are any two finite dimensional Hilbert spaces of the same dimension.*

11 Closed subspaces

The notion of a closed subspace is central to the theory of Banach spaces. Ordinarily, this is simply defined to be a closed linear set, that is, a closed set satisfying $ax + by \in B$ for every x, y in B and a, b in \mathbb{R} . In Section 6, however, we saw that in non-compact spaces much can depend on how one treats the associated topological and metric notions. In the situation at hand, three different notions arise:

- (1) A *closed linear set* is a closed subset of B that is further linear.
- (2) A *closed subspace* is a separably closed subset of B that is linear.
- (3) A *located closed subspace* is a closed subspace with a locating function.

By Theorem 5.1, there is no need to consider located closed linear subsets, since these amount to the same things as located closed subspaces. Note that a closed subspace of a Banach (resp. Hilbert) space inherits the norm (resp. inner product) from the larger space, and so can be considered a Banach (resp. Hilbert) space in its own right. Note also that one can equivalently present a closed subspace as the closure of the linear span of a countable sequence of vectors; taking all finite rational linear combinations of the vectors provides the corresponding linear set. We will use this equivalence freely below. In the case of Hilbert spaces, we know by Theorem 10.9 that in RCA_0 we can even take the closed subspace to be given by an orthonormal basis; and by Theorem 10.10 that in WKL_0 every closed subspace has an independent generating sequence.

In reverse mathematics, one usually uses the notion of a closed subspace, as defined above; in constructive mathematics it is common to require that the subspaces under consideration are furthermore located. All these notions are equivalent in the presence of Π_1^1 comprehension. The goal of this section is to indicate some of the relationships that hold in weaker subsystems of second-order arithmetic.

In fact, all three notions above come up naturally in practice. For example, if f is a bounded linear functional, its kernel, $\ker f$, is clearly a closed linear set; below we will show that, in RCA_0 , it is also a subspace, and that it is located if and only if the norm of f exists. When we turn to the ergodic theorem, we will consider sets of the form $\{x \mid Tx = x\}$, where T is a bounded linear operator from a Hilbert space to itself. It is easy to show that this is always a closed linear set; below we will see that the assertion that it is separably closed in general is equivalent to (ACA) . In our proof of the ergodic theorem we will also consider the closure of sets of the form $\{Tx - x \mid x \in H\}$; this is always a separably closed set, but we will see that the statement that it is always closed is equivalent to (ACA) .

The following theorem collects some of the results we will ultimately obtain:

Theorem 11.1 (RCA_0) *The following hold with respect to both Hilbert spaces and Banach spaces:*

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

The implications from (ACA) and (Π_1^1-CA) in all four statements, for Banach spaces as well as Hilbert spaces, follow from the more general results for metric spaces, Theorem 5.3 and 5.4. Since every Hilbert space is a Banach space, it suffices to obtain the reversals for the former. For 1 and 2, these are found in Theorem 11.2 and Corollary 15.2, respectively. The reversals for 3 and 4 are consequences of Theorems 13.1 and 13.4, though stronger reversals for individual distances are given by Theorems 12.5 and 12.6. Note that statements 2 and 4 are not sharp; see the discussion in Section 16.

Theorem 11.2 (RCA_0) *The statement that every closed subspace of a Hilbert space is closed implies (ACA) .*

Proof. Let $\exists y \theta(x, y, Z)$ be the complete Σ_1 formula relative to Z discussed in Section 2. Let H be the Hilbert space l^2 with orthonormal basis $\langle e_i \mid i \in \mathbb{N} \rangle$. Let $\langle b_i \mid i \in \mathbb{N} \rangle$ be the sequence

$$b_i = \begin{cases} e_j & \text{if } i = \langle j, k \rangle \text{ and } \theta(j, k, Z) \\ 0 & \text{otherwise} \end{cases}$$

and let C be the closed subspace spanned by the sequence $\langle b_i \rangle$. Assuming C is a closed linear set, its complement is open, and we have

$$\begin{aligned} \exists y \theta(i, y, Z) &\leftrightarrow e_i \in C \\ &\leftrightarrow e_i \notin \overline{C} \end{aligned}$$

showing that $\exists y \theta(i, y, Z)$ is equivalent to a Π_1 formula. $\square \quad \square$

A strengthening of this theorem is contained in Theorem 15.1 below.

12 Distances and projections

If M is a closed linear set or a closed subspace, the notion of the distance of a point x to M , and the notion of a distance function for M , carry over from the case of metric spaces. In the case of Hilbert spaces, we can also define the notion of the projection onto a closed subspace.

Definition 12.1 *Let M be a closed subspace of a Hilbert space H . Let x and y be elements of H , with y in M . If the distance from x to y is less than or equal to the distance from x to any other point 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 .*

As was the case with distances, the notion of being a projection makes sense more generally for any class M that can be defined by a formula in the language of second-order arithmetic. In particular, the definition above makes sense for closed linear subsets M as well. Note that if y is the projection of x on M , then $\|x - y\| = d(x, M)$. Using the parallelogram identity,

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

one can show that in a Hilbert space the projection is unique. For, 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

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

and so these are all equalities. Let $d = \|x - y\|$. The parallelogram identity then implies

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

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

Lemma 12.2 (RCA₀) *Let M be linear. An element y is the projection of x on M if and only if y is in M and $x - y$ is orthogonal to M .*

Proof. The standard argument formalizes without difficulty. Suppose y is the projection of x onto M , but $x - y$ is not orthogonal to M . Let $z \in M$ be 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

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

Since the left-hand side is equal to $\langle x - y, x - y \rangle + 2\langle x - y, az \rangle + \langle az, az \rangle$, we obtain that for every $a \in \mathbb{R}$,

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

Since a can be positive or negative, we can choose an a with sufficiently small absolute value to obtain a contradiction.

Conversely, suppose y is in M with $\langle x - y, z \rangle = 0$ for all $z \in M$. Using the Pythagorean theorem we have that for every $y' \neq y$ in M

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

So y is the closest point in M to x , and hence the projection of x onto M . \square

We will often 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 that “there exists a projection function for M .” The next theorem demonstrates the relationship between distances and projections in the context of a subspace.

Theorem 12.3 (RCA₀) *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 implies 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 $\langle w_m \mid m \in \mathbb{N} \rangle$ a dense sequence of points in M , and suppose $d = d(x, M) = \inf\{d(x, y) \mid y \in M\}$ exists. By the definition of an infimum, we have

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

By Lemma 2.2, there is a sequence of points y_n from $\langle w_m \rangle$ such that for every n ,

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

It suffices to show that the sequence $\langle y_n \rangle$ is Cauchy with an explicit rate of convergence, since if $y = \lim_n y_n$, then clearly $d(x, y) = d = d(x, M)$. Since $\frac{1}{2}(y_n + y_m)$ is in M , we have $d(\frac{1}{2}(y_n + y_m), x) \geq d$. Using the parallelogram identity, we then have

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

Since the last quantity can be made arbitrarily small by requiring n and m to be large, we are done.

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 we define its value at each element of the countable dense subset A of H , as above. We only need to check that P is linear and bounded. For linearity, given $a, b \in \mathbb{Q}$,

$x, y \in A$, and $z \in M$, we have

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

so, by Lemma 12.2,

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

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

$$\|x\|^2 = \langle x, x \rangle = \langle x - Px + Px, x - Px + Px \rangle = \|x - Px\|^2 + \|Px\|^2$$

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

If one replaces “closed subspace” by “closed linear subset” in Theorem 12.3, the situation changes. The uniform case stays the same: the existence of a locating function is equivalent to the existence of a projection function. For an individual point, however, it is not clear whether knowing the distance to the set helps find the projection at all.

Theorem 12.4 (RCA₀) *Suppose M is a closed linear subset of a Hilbert space and x is a point.*

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

Proof. As above, if $P_M x$ exists, then $d(x, M) = \|x - P_M x\|$, proving part 1. Similarly, if P_M exists, this equality provides a locating function, proving one direction of 2.

All that remains is to prove the converse direction of 2. So, suppose M is located in a Hilbert space $H = \hat{A}$. To prove that the projection function exists, it suffices to show that one can define the sequence of values $\langle P_M y \mid y \in A \rangle$. In other words, we need to show how the projection y of x on M can be obtained in RCA₀ uniformly from a locating function for M . The construction is similar to that of Theorem 12.3, but instead of choosing each element y_m from a sequence of points that is dense in M , we choose it from the dense subset A of H , using the locating function to ensure that y_m is close enough to M .

Let $d = d(x, M)$. Note that by the definition of distance, for every n there is a point w in M such that $d \leq \|x - w\| < d + 2^{-(n+1)}$. Then there is a point y' in A such that $\|y' - w\| < 2^{-(n+1)}$. Hence we have $\|x - y'\| \geq \|x - w\| - \|y' - w\| > d - 2^{-(n+1)}$, $\|x - y'\| \leq \|x - w\| + \|y' - w\| < d + 2^{-n}$, and $d(y', M) \leq \|y' - w\| < 2^{-(n+1)}$.

Thus we have

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

By Lemma 2.2 there is a sequence $\langle y_n \mid n \in \mathbb{N} \rangle$ such that each y_n witnesses the matrix above for the corresponding n . It suffices to show that $\langle y_n \rangle$ is Cauchy, since if $\langle y_n \rangle$ converges to y , it is clear that $d(y, M) = 0$, implying $y \in M$; and $\|x - y\| = d$, so that y is necessarily the projection of x on M .

As in the proof of Theorem 12.3 we have

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

so it suffices to show that $\|\tfrac{1}{2}(y_n + y_m) - x\|$ is not too much smaller than d . 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 $\tfrac{1}{2}(w' + w'') \in M$, so $\|\tfrac{1}{2}(w' + w'') - x\| \geq d$. Thus we have

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

The proof then proceeds just as the proof of Theorem 12.3. \square

We do not know how things stand with respect to the converse of 1; the strength of the statement “if $d(x, M)$ exists then so does $P_M x$ ” is left as an open question in Section 16.

Given the close relationship between distances and projections, what does it take to show the existence of either?

Theorem 12.5 (RCA₀) *Each of the following statements is equivalent to (ACA):*

- (1) *For every closed subspace M of a Hilbert space, the projection on M exists.*
- (2) *Every closed subspace of a Hilbert space is located.*
- (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.*

Proof. Note that 3 follows immediately from 1. By Theorem 12.3, 1 and 2 are equivalent, as are 3 and 4, and by Theorem 5.4 we have that (ACA) implies 2. To close the chain it suffices to show that 4 implies (ACA); this follows from Theorem 13.4, or, alternatively, from Corollary 15.4 below. \square

If we replace “closed subspace” by “closed linear subset,” we have the following:

Theorem 12.6 (RCA₀) *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 5.4 that they are implied by $(\Pi_1^I\text{-CA})$. Also, each of 1 and 2 implies 3, which in turn implies 4. The fact that 4 implies (ACA) is given by Corollary 15.2 below. \square

The results of this section can be made more general. The notion of projection does not make sense in arbitrary Banach spaces, because there need not be a unique closest point to a given point x in a closed subspace M . This is, however, always the case for *uniformly convex Banach spaces*; these are Banach spaces with the additional property that for every two sequences of vectors $\langle x_n \mid n \in \mathbb{N} \rangle$ and $\langle y_n \mid n \in \mathbb{N} \rangle$, if $\|x_n\| \rightarrow 1$, $\|y_n\| \rightarrow 1$, and $\|x_n + y_n\| \rightarrow 2$, then $\|x_n - y_n\| \rightarrow 0$. In subsystems of second-order arithmetic, it is reasonable to ask that such spaces come equipped with a *modulus of convexity*, that is, a function $\delta(\varepsilon)$ from \mathbb{Q} to \mathbb{Q} such that whenever $\|u\| \leq 1$, $\|v\| \leq 1$, and $\frac{1}{2}\|u - v\| \geq 1 - \delta(\varepsilon)$, then $\|u - v\| \leq \varepsilon$. For such spaces, the analogs of Theorems 12.3–12.5 are provable in RCA_0 . The modulus of convexity is used in place of the parallelogram identity, e.g. to show that the sequence $\langle y_n \rangle$ defined in the proof of Theorem 12.3 is a Cauchy sequence.

13 Norms and kernels of bounded linear functionals

Recall that by a bounded linear functional we mean a bounded linear operator from the space in question to \mathbb{R} , and, in RCA_0 , every bounded linear functional on a Banach space is equivalent to a linear continuous function. As a result, if f is a bounded linear functional, it is clear that $\ker f$ is a closed linear set, since it is the inverse image of the closed set $\{0\}$. In this section we will show that it is also a closed subspace, provably in RCA_0 , and that it is located if and only if the norm of f exists.

Theorem 13.1 (RCA_0) *The kernel of any bounded linear functional on a Banach space is a closed subspace as well as a closed linear set.*

Proof. We have already noted that $\ker f$ is a closed linear set. If f is the constant zero function, the statement that $\ker f$ is a closed subspace is trivial. Thus we can assume that there is a $y \in B$ such that $f(y) \neq 0$. Replacing y by $y/f(y)$ if necessary, we can assume that $f(y) = 1$.

Let M be a bound on the norm of f , so for every x , $|f(x)| \leq M\|x\|$. In particular, we have $M\|y\| \geq |f(y)| = 1$. Let $A = \langle a_n \mid n \in \mathbb{N} \rangle$ be a sequence of points that is dense in B , and define the sequence $\langle x_n \mid n \in \mathbb{N} \rangle$ by $x_n = a_n - f(a_n)y$. Then for each n we have

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

so x_n is in the kernel of f . It suffices to show that the sequence x_n is dense in the kernel, that is, for every $\varepsilon > 0$ and every x such that $f(x) = 0$, there is a point $a \in A$ such that $\|x - (a - f(a)y)\| < \varepsilon$.

Because A is dense in B , we may choose $a \in A$ such that $\|x - a\| < \varepsilon/(2M\|y\|)$. By the choice of M we then have $\|x - a\| < \varepsilon/2$, as well as $|f(a)|\|y\| = |f(x - a)|\|y\| \leq M\|x - a\|\|y\| \leq \varepsilon/2$. Therefore

$$\|x - (a - f(a)y)\| \leq \|x - a\| + |f(a)|\|y\| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

as required. \square

When is the kernel of a functional located? The following theorem is adapted from [1].

Theorem 13.2 (RCA₀) *Let f be any bounded linear functional on a Banach space. The following statements are equivalent:*

- (1) *The norm of f exists.*
- (2) *$\ker f$ is located.*

Proof. 1 \rightarrow 2: Let f be a bounded linear functional with norm $\|f\|$. If $f \equiv 0$ then $d(x, \ker f) = 0$ for all x , so assume $\|f\| > 0$. We will show that for every $x \in B$,

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

This implies that $d(x, \ker f)$ is the desired locating function.

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

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

and so $\|x - z\| \geq |f(x)|/\|f\|$. This implies $d(x, \ker f) \geq |f(x)|/\|f\|$.

In the other direction, we need to show that $d(x, \ker f) \leq |f(x)|/\|f\|$. Since $\|f\| > 0$, there is an $\varepsilon' > 0$ such that $\|f\| > \varepsilon'$. We have

$$\|f\| = \sup\left\{\frac{|f(x)|}{\|x\|} \mid x \in H, x \neq 0\right\}$$

So for any $\varepsilon < \varepsilon'$, there exists $y \in H$ 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 ε can be made arbitrarily small, we have $d(x, \ker f) \leq |f(x)|/\|f\|$, as required.

As for $2 \rightarrow 1$, suppose $\ker f$ is located. Again, if f is identically zero, $\|f\| = 0$. Otherwise, there is an x_0 such that $f(x_0) = 1$, and, since f is continuous, $d(x_0, \ker f) > 0$. Then we have

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

The first equality follows from the definition of distance. The second equality follows from the fact that if $y = x_0 - z$ then $f(y) = f(x_0) - f(z) = 1$; and, on the other hand, if $f(y) = 1$, we can write $y = x_0 - (y - x_0)$, with $y - x_0 \in \ker f$. Now, if $f(y) = 1$ and $x = y/\|y\|$, then $\|x\| = 1$ and $f(x) = 1/\|y\|$; and if $\|x\| = 1$ and $y = x/f(x)$, then $f(y) = 1$ and $\|y\| = 1/|f(x)|$. Thus we have

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

showing that $\|f\|$ exists. \square

For Hilbert spaces, we have the following refinement. Note that clause 4 is one form of the Riesz Representation Theorem.

Theorem 13.3 (RCA₀) *Let f be any bounded linear functional on a Hilbert space H . The following are equivalent:*

- (1) *The norm of f exists.*
- (2) *$\ker f$ is located.*
- (3) *If f is not identically 0, then $d(x, \ker f)$ exists for some x such that $f(x) \neq 0$.*
- (4) *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, and clearly 2 implies 3. It remains to show $3 \rightarrow 4$ and $4 \rightarrow 1$.

$3 \rightarrow 4$: If $f \equiv 0$, we can simply take $y = 0$. Otherwise, suppose $f(x_0) \neq 0$ and $d(x_0, \ker f)$ exists. By Theorem 13.1, $\ker f$ is a closed subspace, and by

Theorem 12.3, the projection Px_0 of x_0 on $\ker f$ exists. Let $z = x_0 - Px_0$, so, by Lemma 12.2, z is orthogonal to $\ker f$. Let $w = z/\|z\|$, and let $y = wf(w)$. Then $\|w\| = 1$, $\|y\| = |f(w)|$, $f(y) = (f(w))^2$, and y is still orthogonal to $\ker f$. Then for any x ,

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

so $x - (f(x)/f(y))y \in \ker f$. Then

$$\begin{aligned} 0 &= \left\langle x - \frac{f(x)}{f(y)}y, y \right\rangle \\ &= \langle x, y \rangle - \frac{f(x)}{f(y)}\|y\|^2 \\ &= \langle x, y \rangle - f(x). \end{aligned}$$

This implies $f(x) = \langle x, y \rangle$, as required.

4 \rightarrow 1. Suppose $f(x) = \langle x, y \rangle$ for every x . Then $f(y) = \|y\|^2$, so $\|f\| \geq \|y\|$. On the other hand, by the Cauchy-Schwartz inequality, $|f(y)| = |\langle y, y \rangle| \leq \|y\|^2$, and so $\|f\| \leq \|y\|$. \square

Theorem 13.4 (RCA₀) *The following statements are equivalent to (ACA):*

- (1) *Every bounded linear functional on a Banach space has a norm.*
- (2) *Every bounded linear functional on a Hilbert space has a norm.*
- (3) *For every bounded linear functional f on a Banach space, $\ker f$ is located.*
- (4) *For every bounded linear functional f on a Hilbert space, $\ker f$ is located.*
- (5) *For every bounded linear functional f on a Hilbert space, if f is not identically 0, then $d(x, \ker f)$ exists for some x such that $f(x) \neq 0$.*
- (6) *Every bounded linear functional f on a Hilbert space is representable, i.e. there is a y in the space such that, for every x , $f(x) = \langle x, y \rangle$.*

Proof. By the preceding theorems, 1 and 3 are equivalent, as are 2, 4, 5, and 6. Since clearly 1 implies 2, it suffices to show (ACA) implies 1 and 6 implies (ACA).

(ACA) \rightarrow 1: If f is not identically 0, then $\|f\| = \sup\{|f(x)|/\|x\| \mid 0 \neq x \in A\}$. Since f is a bounded linear functional, $\{|f(x)|/\|x\| \mid 0 \neq x \in A\}$ is a bounded sequence of real numbers, and (ACA) implies that it has a least upper bound.

6 \rightarrow (ACA). We will use Lemma 2.1. Let $\langle b_n \mid n \in \mathbb{N} \rangle$ be any increasing sequence of real numbers in $[0, 1]$ such that for each n , $\sum_{i < n} b_i^2 \leq 1$. We will show that 6 implies that $\sum b_i^2$ exists. Let H be l^2 , the collection of all square summable countable sequences of reals, with orthonormal basis $\langle e_n \mid n \in \mathbb{N} \rangle$. Define a functional f on the orthonormal basis by $f(e_i) = b_i$ and extend it

linearly to all of H . To show f is bounded, we use Hölder's inequality, which can be formalized in RCA_0 : for each n ,

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

So the norm of f is bounded by 1.

Assuming 6, we have that $f(x) = \langle x, y \rangle$ for some y , in which case $\|f\| = \|y\|$. From (2), we see that $\|y\| = \|f\| \leq (\sum_i b_i^2)^{\frac{1}{2}}$, assuming the latter exists. On the other hand, if we take $x_i = b_i$ for $i < n$ in (2), we have $\sum_{i < n} f(x_i e_i) = \sum_{i < n} b_i^2$, and so for each n , $\|y\| = \|f\| \geq (\sum_{i < n} b_i^2)^{\frac{1}{2}}$. Hence $\|y\| = \|f\| = (\sum_i b_i^2)^{\frac{1}{2}}$, and so $\|y\|^2 = \sum b_i^2$ exists, as required. \square

14 The mean ergodic theorem

Although von Neumann's *mean ergodic theorem* (see, for example, [19,25]) was initially stated in the context of a measure space, it can be stated and proved naturally in the more general context of a Hilbert space. In this setting, the theorem is as follows:

If T is an isometry of a Hilbert space and x is any point, then the sequence $\langle S_n x \mid n \in \mathbb{N} \rangle$ of partial averages

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

converges.

Here an *isometry* is a linear transformation T satisfying $\|Tx\| = \|x\|$ for every x . The theorem holds more generally if T is any *nonexpansive* linear transformation, i.e. T satisfies $\|Tx\| \leq \|x\|$ for every x . Think of x as describing some measurement on a physical system, depending on the system's state, and think of Tx as denoting the same measurement taken after a unit of time. The mean ergodic theorem states that sequence of partial averages converges in the Hilbert space norm.

(The mean ergodic theorem originally dealt with measure preserving transformations U on a measure space X . Such a transformation gives rise to an isometry T of the Hilbert space $L^2(X)$ defined by $(Tf)(x) = f(U(x))$. In this case, the mean ergodic theorem asserts that the sequence of time averages, as

a function of the initial state, converges in the L^2 norm. The Birkhoff *pointwise ergodic theorem* implies that the sequence converges pointwise almost everywhere, and in the L^1 norm. We will not consider the pointwise ergodic theorem here.)

In fact, the standard proof of the mean ergodic theorem gives more information. 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 $\{Tx - x \mid x \in H\}$. Then M and N are closed subspaces, and standard proof shows:

M and N are orthogonal complements to one another, and the sequence $\langle S_n x \rangle$ converges to the projection of x on M .

We need to do some work to make sense of this in the context of subsystems of second-order arithmetic. Given an isometry or nonexpansive mapping T , note that M and N can certainly be described by formulas in the language of second-order arithmetic, so the notion of a projection makes sense. In fact, it is not hard to see that M is a closed linear set, since it is the kernel of the continuous function $f(x) = \|Tx - x\|$; and that N is a closed subspace, since the set $\{Tx - x \mid x \in A\}$ forms a countable dense subset.

Note that if T is a nonexpansive mapping, then for every x and y we have

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

So, if T is considered as a continuous function, the identity is a modulus of uniform continuity for T . Using Lemma 2.2 we can therefore make sense of the sequence of iterations T^n , and hence also the partial averages S_n . Since, in general, the statement that T is linear (resp. an isometry, nonexpansive operator) is Π_1 , we can show by induction in RCA_0 that if T is an isometry (resp. a nonexpansive linear operator) then so is T^n for each n . Similarly, we can show that each S_n exists, and is nonexpansive.

It turns out that in either formulation the mean ergodic theorem is equivalent to arithmetic comprehension:

Theorem 14.1 (RCA_0) *Each of the following statements is equivalent to (ACA):*

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

Clearly 1 implies 2. The fact that 2 implies (ACA) is a consequence of Theorem 15.3 below, and the fact that (ACA) implies 1 is Corollary 14.5 below.

In fact, our analysis will yield much more information. The following theorem, essentially from Spitters [23, Section 7.2], spells out some of the relationships between existence statements that are implicit in conventional proofs of the mean ergodic theorem. Our proof follows that of [23], but our Lemma 14.4 is new and allows us to avoid an appeal to the Riesz representation theorem.

Theorem 14.2 (RCA₀) *Let T be any nonexpansive linear operator on a Hilbert space, let x be any point. With the notation above, the following are equivalent:*

- (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) $\lim_{n \rightarrow \infty} S_n x$ exists.

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

$$\lim_n S_n x = P_M x = x_M = x - P_N x$$

To prove this, we isolate two helpful lemmas.

Lemma 14.3 (RCA₀) *Let T be any nonexpansive linear operator on a Hilbert space, and let N be the closure of the set $\{Tx - x \mid x \in H\}$. Then for every x in N , $\lim_n S_n x = 0$.*

Proof. Suppose $x \in N$. Then for every $\varepsilon > 0$ there is a $u \in H$ such that $\|x - (u - Tu)\| < \varepsilon$. Then we have

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

and so

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

Also, since $\|x - (u - Tu)\| < \varepsilon$, we have

$$\|S_n x - S_n(u - Tu)\| = \|S_n(x - (u - Tu))\| < \varepsilon.$$

Since ε was arbitrary, we have $S_n x \rightarrow 0$ as well. \square

Lemma 14.4 (RCA₀) *Let T be any nonexpansive linear operator on a Hilbert space, let $M = \{x \mid Tx = x\}$ and let N be the closure of the set $\{Tx - x \mid x \in H\}$. Then $N^\perp = M$, that is, every element orthogonal to N is in M and vice-versa. Hence, $M \cap N = \{0\}$.*

Proof. Suppose $x \in H$ is orthogonal to N , i.e. $\langle x, y - Ty \rangle = 0$ for all $y \in H$. In particular, $\langle x, x - Tx \rangle = 0$, or

$$\langle x, Tx \rangle = \langle x, x \rangle = \|x\|^2$$

So

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

Therefore, $x = Tx$, and so x is in M , as required.

Conversely, suppose x is in M , i.e. $Tx = x$. To show that x is orthogonal to N , it suffices to show that $\langle x, Ty - y \rangle = 0$ for every $y \in H$. Since T is nonexpansive, for every $\alpha \geq 0$ we have

$$\|x + \alpha Ty\| = \|T(x + \alpha y)\| \leq \|x + \alpha y\|,$$

and so, squaring both sides, we have

$$\|x\|^2 + 2\alpha\langle x, Ty \rangle + \alpha^2\|Ty\|^2 \leq \|x\|^2 + 2\alpha\langle x, y \rangle + \alpha^2\|y\|^2.$$

Hence

$$\langle x, Ty - y \rangle \leq \frac{\alpha}{2}(\|y\|^2 - \|Ty\|^2)$$

for every α , so $\langle x, Ty - y \rangle \leq 0$. For every $\alpha \leq 0$, the same calculation shows

$$\langle x, Ty - y \rangle \geq \frac{\alpha}{2}(\|Ty\|^2 - \|y\|^2).$$

So $\langle x, Ty - y \rangle = 0$, as required.

To see that $M \cap N = \{0\}$, note that if x is in both M and N , then $\langle x, x \rangle = 0$, so $x = 0$. \square

We now turn to the proof of Theorem 14.2.

Proof. 1 \rightarrow 2: Write $x = (x - P_N x) + P_N x$. Then $P_N x$ is in N , and since $x - P_N x$ is orthogonal to N , the previous lemma implies $x - P_N x$ is in M .

2 \rightarrow 3: If $x \in H$ and $x = x_M + x_N$ then $S_n(x_M) = x_M$ for all n , so $\lim_n S_n(x_M) = x_M$. Lemma 14.3 implies $\lim_n S_n(x_N) = 0$. So $\lim_n S_n x = x_M$.

3 \rightarrow 1: Let $y = \lim S_n x$, and let $z = x - y$. We will show that z is the projection of x on N . By Lemma 12.2, it suffices to show that z is in N and $x - z$ is orthogonal to N .

To see that z is in N , for each n let

$$y_n = \left(\frac{n-1}{n}I + \frac{n-2}{n}T + \dots + \frac{1}{n}T^{n-1} \right)x.$$

Then for every n , $y_n - Ty_n = \frac{n-1}{n}x - S_nx \rightarrow x - y = z$, so z is in N .

To see that $x - z = y$ is orthogonal to N , by Lemma 14.4 it suffices to show that y is in M , i.e. $Ty = y$. This follows from the fact that

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

and so

$$Ty = \lim_n TS_nx = \lim_n S_nx = y,$$

as required.

To see that the decomposition in statement 2 is unique, suppose $x_M + x_N = x'_M + x'_N$. Then $x_M - x'_M = x'_N - x_N$ is in $M \cap N$, and Lemma 14.4 applies these differences are equal to 0. The proof of $3 \rightarrow 1$ establishes that $P_Nx = x - \lim_n S_nx$. It also establishes that $\lim_n S_nx$ is in M , and so to show that $P_Mx = \lim_n S_nx$, it suffices to show that $x - \lim_n S_nx$ is orthogonal to M . But this also follows from Lemma 14.4: if y is any element of M , it is orthogonal to every element of N , including $x - \lim_n S_nx$. \square

Corollary 14.5 (ACA₀) *The mean ergodic theorem holds, i.e. for every H and nonexpansive mapping T , for every $x \in H$, S_nx converges. In fact, both P_M and P_N exist, and for every x , $P_Mx = x - P_Nx = \lim_n S_nx$.*

Proof. Since N is a closed subspace, ACA₀ proves that P_N exists by Theorem 12.3. Theorem 14.2 then finishes it off. \square

Note that the statement “ P_Mx exists” is notably absent from the list of equivalent statements in Theorem 14.2. This was not an oversight: in fact, Theorem 15.3 below shows that the hypothesis that P_Mx exists provides no help at all in proving the mean ergodic theorem! That is, on the assumption that P_Mx exists, we can show that if $\lim_n S_nx$ exists it *has* to be equal to P_Mx ; but even with the assumption that P_Mx exists and is equal to 0, it still requires (ACA) to show that $\lim_n S_nx$ exists. (Similarly, although, by Lemma 14.4, RCA₀ proves $N^\perp = M$, in general it requires (ACA) to show that $M^\perp = N$; the sticking point is showing $M^{\perp\perp} \subseteq M$.)

15 Complex Hilbert spaces and the remaining reversals

The mean ergodic theorem asserts the existence of something characterizing the long term behavior of a system evolving over time. One would not expect this theorem to be provable in RCA_0 , since that would imply that such limiting behavior can be determined computably. For example, suppose $\langle a_i \mid i \in \mathbb{N} \rangle$ is a sequence of reals in $[0, 1]$, and define an operator T on l^2 by $Te_i = (1 - a_i)e_i$. Then T is a nonexpansive mapping, and it is not hard to see that $S_n e_i$ remains fixed at e_i if $a_i = 0$, and converges to 0 otherwise. Let $x = \sum_i 2^{-i} e_i$, 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 Σ_1 equivalent to the Π_1 assertion $a_i = 0$. By Lemma 2.1, this shows that the mean ergodic theorem for nonexpansive mappings implies arithmetic comprehension.

To extend the reversal for the mean ergodic theorem to *isometries*, it is helpful to introduce the notion of a Hilbert space over the complex numbers. These can be defined in subsystems of second-order arithmetic in the expected way, replacing \mathbb{R} by \mathbb{C} in Definition 9.3 and changing clause 2 to $\langle x, y \rangle = \overline{\langle y, x \rangle}$. The notion of a bounded linear operator can be lifted accordingly. Note that any n -dimensional Hilbert space over the complex numbers can be viewed as a $2n$ -dimensional Hilbert space over the real numbers: if $\{e_0, \dots, e_{n-1}\}$ is an orthonormal basis for a complex space, a vector $\sum_j u_j e_j$ with complex coefficients u_j can alternatively be viewed as a vector $\sum_j (\text{Re}(u_j)e'_{2j} + \text{Im}(u_j)e'_{2j+1})$ in a real Hilbert space with orthonormal basis $\{e'_0, e'_1, \dots, e'_{2n-2}, e'_{2n-1}\}$. Similarly, any infinite-dimensional complex Hilbert space can be viewed as an infinite-dimensional real Hilbert space, and any bounded linear operator in the sense of the complex space can be viewed as a bounded linear operator in the sense of the real space, with the same norm.

Given any sequence $\langle a_k \mid k \in \mathbb{N} \rangle$ of real numbers in $[0, 1]$, define the sequence of complex numbers

$$\beta_k = \frac{1 + a_k i}{|1 + a_k i|}.$$

So $|\beta_k| = 1$ for each k and $\beta_k = 1$ if and only if $a_k = 0$. Define a linear operator T on the complex Hilbert space l^2 of square-summable sequences of complex numbers by $Te_k = \beta_k e_k$. The fact that $|\beta_k| = 1$ for each k implies that T is an isometry. Once again, if $\beta_k = 1$, then $S_n e_k$ is fixed at e_k ; otherwise,

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

which converges to 0 as n increases, since

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

As above, if $x = \sum 2^{-k} e_k$ and $y = \lim_n S_n x$, then we have $\langle y, e_k \rangle \neq 0$ if and only if $\beta_k = 1$, i.e. if and only if $a_k = 0$. Once again, by Lemma 2.1, this implies arithmetic comprehension. Readers who prefer to think of the mean ergodic theorem in terms of measure-preserving transformations can recast this argument in terms of a measure preserving operator U on ω copies of the torus $\mathbb{R}/(2\pi\mathbb{Z})$, where U rotates the k th component a small amount if $a_k > 0$ and leaves it fixed otherwise.

The sketch above provides one way of completing the proof of Theorem 14.1. In fact, Theorem 15.3 provides an even stronger result. But first, we can use these constructions to pay off some old debts. Remember that for $M = \{x \mid Tx = x\}$ and N the closure of $\{Tx - x \mid x \in H\}$, as in the proof of the mean ergodic theorem, RCA_0 proves that M is a closed linear set, and N is a closed subspace.

Theorem 15.1 (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 14.5, (ACA) implies that both P_M and P_N exist for any nonexpansive mapping T . This implies 5 right away. Also, by Theorems 12.4 and 12.3 respectively, it implies that M and N are both located; and by Theorem 5.1 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 of Lemma 2.1. 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 theorem. 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 Σ_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 Σ_1 definition of $\{k \mid a_k = 0\}$.

To show 8 implies (ACA), once again we use the same construction. Let $x = \sum 2^{-k} e_k$. Assuming 8, $d = d(x, M)$ exists. Define a sequence $\langle c_n \mid n \in \mathbb{N} \rangle$ recursively, as follows:

$$c_n = \begin{cases} 0 & \text{if } d^2 - \sum_{k < n} c_k \cdot 4^{-k} < \frac{1}{2} \cdot 4^{-n} \\ 1 & \text{if } d^2 - \sum_{k < n} c_k \cdot 4^{-k} > \frac{1}{2} \cdot 4^{-n} \\ \text{either 0 or 1 otherwise.} \end{cases}$$

We can do this because the relation $x \leq y$ is Π_1 , and for every x and y , either $x \leq y$ or $y \leq x$; thus c_n is obtained by searching for a witness to the failure of one condition or the other. Using induction on n , we can show

$$\forall k < n (c_k = 1 \leftrightarrow a_k \neq 0),$$

as follows. Assuming the statement is true for n , we need to show that $c_n = 1 \leftrightarrow a_n \neq 0$. Suppose $a_n \neq 0$. By the inductive hypothesis, we have that for every $k < n$, whenever $c_k = 1$, then $a_k \neq 0$; this in turn implies $e_k \notin M$. Similarly, by assumption, $a_n \neq 0$ and so $e_n \notin M$. Thus if $\sum \alpha_k e_k$ is any element of M , we have $\forall k < n (c_k = 1 \rightarrow \alpha_k = 0)$, and $\alpha_n = 0$. So

$$d(x, M)^2 \geq \sum_{k < n} c_k \cdot 4^{-k} + 4^{-n},$$

which implies $c_n = 1$.

On the other hand, suppose $a_n = 0$. By the inductive hypothesis, we have that for every $k < n$, if $c_k = 0$ then $a_k = 0$, which, in turn, implies $2^{-k} e_k \in M$. By

our assumption, we also have $2^{-n}e_n \in M$. So

$$\begin{aligned} d(x, M)^2 &\leq \sum_{k < n} c_k \cdot 4^{-k} + \sum_{k > n} 4^{-k} \\ &= \sum_{k < n} c_k \cdot 4^{-k} + \frac{1}{3} \cdot 4^{-n}, \end{aligned}$$

whence $c_n = 0$. Thus $\langle c_n \rangle$ works as advertised, and so $S = \{n \mid a_n \neq 0\} = \{n \mid c_n = 1\}$ exists. \square

Corollary 15.2 (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 first implication duplicates the conclusion of Theorem 11.2, whereas the second completes the proof of Theorem 11.1. The third and fourth implications complete the proof of Theorem 12.6.

Theorem 15.3 RCA_0 *proves that the following are equivalent to (ACA):*

- (1) *For every isometry T on a Hilbert space and any point x , $\lim_n S_n x$ exists.*
- (2) *For every isometry T on a Hilbert space and any point x , if $P_M x$ exists, then $\lim_n S_n x$ exists.*
- (3) *For every isometry T on a Hilbert space and any point x , if $P_M x = 0$, then $\lim_n S_n x$ exists.*
- (4) *For every isometry T on a Hilbert space and any point x , if $P_M x = 0$, then $\lim_n S_n x = 0$.*

Corollary 14.5 implies (ACA) implies 1. Clearly 1 implies 2 and 2 implies 3. The fact that 3 implies 4 is given by Theorem 14.2. So, we only need to show that 4 implies (ACA).

By the observations above, nothing is lost if we interpret statement 4 in terms of complex Hilbert spaces. The existential content of statement 4 becomes clearer if we write it in its contrapositive form:

For every isometry T on a Hilbert space and any point x , if $\lim_n S_n x$ either fails to exist or is not equal to 0, then $P_M x$ either fails to exist or is not equal to 0.

Note that the conclusion implies, in particular, there exists a element y of

M such that $\langle x, y \rangle \neq 0$. To obtain the reversal, then, it suffices to prove the following in RCA_0 : given a sequence $\langle a_k \mid k \in \mathbb{N} \rangle$ in $[0, 1]$ such that the partial sums $\sum_{k \leq n} a_k^2$ are all bounded by 1, there is a complex Hilbert space H , a point x , and an isometry T such that

- the sequence $\lim_n S_n x$ is bounded away from 0; and
- the existence of any y satisfying $Ty = y$ and $\langle x, y \rangle \neq 0$ implies $\sum a_k^2$ exists.

To obtain such a construction, we will use a strategy employed in a different context by [20, Section 4.4]. Let e_0, e_1, e_2, \dots denote the standard basis on l^2 . We will describe another basis f, e_1, e_2, \dots that cannot, in general, be computably obtained from the first, or vice-versa. We will carry out the construction by *thinking* in terms of the first basis but *proceeding formally* in terms of the second.

This paragraph and the next, then, are purely heuristic. First, by shifting the sequence $\langle a_k \rangle$ if necessary, we can assume without loss of generality that $a_0 = 0$; multiplying each term e.g. by $1/2$ we can also assume that the partial sums $\sum_{k < n} a_k^2$ are bounded strictly below 1. Let

$$f = \gamma e_0 + \sum_{k \geq 1} a_k e_k$$

where γ is chosen so that $\|f\| = 1$; in other words,

$$\gamma = \sqrt{1 - \sum a_k^2}.$$

Clearly the existence of γ is equivalent to the existence of $\sum a_k^2$. (In the actual proof below, we will show how to define a Hilbert space in terms of the basis f, e_1, e_2, \dots , without assuming the existence of γ .) Let T be the isometry defined by

$$\begin{aligned} T e_0 &= e_0 \\ T e_k &= \frac{1 - i/2^k}{\|1 - i/2^k\|} e_k \quad \text{for } n \geq 1, \end{aligned}$$

so that, as n increases, $S_n e_0$ stays fixed at e_0 , while $S_n e_k$ converges to 0 for each $k \geq 1$. It should be clear, then, that as n increases $S_n f$ approaches γe_0 . In other words, the existence of $\lim_n S_n f$ implies the existence of γ .

To carry out the reversal we will show the following, in RCA_0 :

- One can define the Hilbert space and operator T above, with respect to the basis f, e_1, e_2, \dots , solely in terms of the sequence $\langle a_k \rangle$;
- The fact that $\sum a_k^2$ is bounded strictly below 1 is enough to guarantee that it is not the case that $\lim_n S_n f = 0$;

- The statement that $P_M f \neq 0$ implies the existence of γ , and hence of $\sum a_k^2$.

The real proof that 4 implies (ACA) now follows.

Proof. Let $\langle a_k \mid k \in \mathbb{N} \rangle$ be any sequence of elements of $[0, 1]$ as above, that is, such that $a_0 = 0$ and the partial sums $\sum_{k < n} a_k^2$ are strictly bounded by $1/2$. Assuming 4, we will show that $\sum a_k^2$ exists.

We define the (complex) Hilbert space H in terms of a basis f, e_1, e_2, \dots by specifying the value of the inner product on these basis elements:

$$\begin{aligned} \langle e_k, e_j \rangle &= \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases} \\ \langle f, f \rangle &= 1 \\ \langle f, e_k \rangle &= \langle e_k, f \rangle = a_k. \end{aligned}$$

Note that although the norm of each basis element is 1, the last clause means that in interesting cases the basis will not be orthogonal; although Theorem 10.9 implies the existence of an orthonormal basis, we will have no use for it. In a moment we will see that the basis vectors are linearly independent. The inner product extends linearly to the set of linear combinations of the basis elements with coefficients from $\mathbb{Q}(i)$, and then to the completion of this set, H .

We claim that every element of the Hilbert space has a unique representation as a sum of the form

$$\alpha_0 f + \sum_{k \geq 1} \alpha_k e_k.$$

To prove existence, let x be any element of the Hilbert space. By definition, x is the limit of a Cauchy sequence $\langle c_n \mid n \in \mathbb{N} \rangle$ with an explicit rate of convergence, where each c_n is a finite linear combination

$$c_n = \beta_{n,0} f + \sum_{1 \leq k < m_n} \beta_{n,k} e_k.$$

We claim first that the sequence $\langle \beta_{n,0} \mid n \in \mathbb{N} \rangle$ is a Cauchy sequence with an explicit rate of convergence. Considering differences between the c_k 's, it suffices to show in general that if $\gamma_0 f + \sum_{1 \leq k < m} \gamma_k e_k$ is any finite linear combination of f and e_1, \dots, e_{m-1} , and

$$\|\gamma_0 f + \sum_{1 \leq k < m} \gamma_k e_k\| < \varepsilon,$$

then $|\gamma_0| < 2\varepsilon$. Since e_1, \dots, e_{m-1} are orthonormal, it is easy to check that $\gamma_0 f - \gamma_0 \sum_{1 \leq k < m} \langle f, e_k \rangle e_k$ is orthogonal to the subspace spanned by e_1, \dots, e_k .

Hence, by Lemma 12.2, $\gamma_0 \sum_{1 \leq k < m} \langle f, e_k \rangle e_k$ is the projection of $\gamma_0 f$ onto this subspace. So we have

$$\begin{aligned}
\varepsilon &> \left\| \gamma_0 f + \sum_{1 \leq k < m} \gamma_k e_k \right\| \\
&\geq \left\| \gamma_0 f - \gamma_0 \sum_{1 \leq k < m} \langle f, e_k \rangle e_k \right\| \\
&\geq |\gamma_0| \left(\|f\| - \left\| \sum_{1 \leq k < m} a_k e_k \right\| \right) \\
&= |\gamma_0| \left(1 - \sum_{1 \leq k < m} a_k^2 \right) \\
&\geq |\gamma_0|/2,
\end{aligned}$$

so $|\gamma_0| < 2\varepsilon$, as required. Thus the sequence $\beta_{n,0}$ converges to a complex number α_0 . Now define a sequence

$$c'_n = \sum_{1 \leq k < m_n} \beta_{n,k} e_k,$$

by deleting the first term of each c_n . It is easy to check that the resulting sequence is again a Cauchy sequence with an explicit rate of convergence. Since it lies entirely in the subspace spanned by the orthonormal basis e_1, e_2, \dots its limit has a representation of the form $\sum_{k \geq 1} \alpha_k e_k$, whence

$$x = \lim_n c_n = \alpha_0 f + \sum_{k \geq 1} \alpha_k e_k.$$

To show uniqueness, it suffices to show that if $\alpha_0 f + \sum_{k \geq 1} \alpha_k e_k = 0$, then $\alpha_k = 0$ for every k . An argument similar to the one above shows first that the assumption implies $\alpha_0 = 0$, and then taking inner products with each e_k shows that each $\alpha_k = 0$.

Next, we define the transformation T . We do this by defining its behavior on the basis vectors: let $T e_n = \beta_n e_n$ for $n \geq 1$, where

$$\beta_n = \frac{1 + i/2^n}{|1 + i/2^n|},$$

and let

$$T f = f + \sum (\beta_n - 1) a_n e_n.$$

The sum in the definition of $T f$ is convergent because for each n , $\|(\beta_n - 1) a_n e_n\| \leq 2^{-n}$. We then extend T linearly to the entire space. To show that T is an isometry, it suffices to show that T preserves inner products on the

basis elements. This can be done by straightforward calculation; for example,

$$\begin{aligned}
\langle Tf, Tf \rangle &= \langle f + \sum (\beta_k - 1)a_k e_k, f + \sum (\beta_k - 1)a_k e_k \rangle \\
&= \langle f, f \rangle + \sum (\overline{\beta_k - 1})a_k^2 + \sum (\beta_k - 1)a_k^2 + \sum (\beta_k - 1)(\overline{\beta_k - 1})a_k^2 \\
&= \langle f, f \rangle + \sum 2(\operatorname{Re}(\beta_k) - 1)a_k^2 + \sum 2(1 - \operatorname{Re}(\beta_k))a_k^2 \\
&= \langle f, f \rangle.
\end{aligned}$$

By induction on n we have

$$T^n f = f + \sum_k (\beta_k^n - 1)a_k e_k$$

and

$$S_n f = f + \sum_k \gamma_{k,n} a_k e_k,$$

where $\gamma_{k,n} = (1 - \beta_k^n)/(n(1 - \beta_k)) - 1$. We have $\lim_n \gamma_{k,n} = -1$. So, if $\lim_n S_n f$ exists, it has to be $f - \sum_k a_k e_k$; in particular, by the assumption on $\langle a_n \rangle$, we know

$$\|f - \sum_k a_k e_k\| \geq \|f\| - \sum_k a_k^2 > 0,$$

so $\lim_n S_n f \neq 0$.

Assuming 4, then, $P_M f \neq 0$. In particular, there is a y such that $\langle f, y \rangle$ is not zero and $Ty = y$. Let $y = \alpha_0 f + \sum_k \alpha_k e_k$. Then, from the definition of T ,

$$Ty = \alpha_0(f + \sum_k (\beta_k - 1)a_k e_k) + \sum_k \alpha_k \beta_k e_k.$$

$Ty = y$ means that for all k ,

$$\alpha_0 \beta_k a_k - \alpha_0 a_k + \alpha_k \beta_k - \alpha_k = (\beta_k - 1)(\alpha_0 a_k - \alpha_k) = 0,$$

so $\alpha_k = -\alpha_0 a_k$ for every k . Then

$$y = \alpha_0(f - \sum_k a_k e_k).$$

Since y is nonzero we know that α_0 is nonzero, so we can define the element y' by

$$y' = f - \frac{y}{\alpha_0} = \sum_k a_k e_k.$$

But then $\sum a_k^2 = \|y'\|^2$ exists, as required. \square

Corollary 15.4 (RCA₀) *Each of the following is equivalent to (ACA):*

- (1) *If S is any closed subspace of a Hilbert space and x is any point, then the distance from x to S exists.*

- (2) If S is any closed subspace of a Hilbert space and x is any point, then the projection of x on S exists.
- (3) If T is any nonexpansive mapping on a Hilbert space H , x is any point, and N is the closure of $\{Tx - x \mid x \in H\}$, then P_Nx exists.
- (4) If T is any isometry on a Hilbert space and x is any point, then P_Nx exists.

Proof. By Theorem 12.3 statements 1 and 2 are equivalent, and by Theorem 12.5 they are provable from (ACA). Clearly 2 implies 3 and 3 implies 4. The fact that 4 implies (ACA) follows from the previous theorem, noting that, by Theorem 14.2, the existence of P_Nx implies the the existence of $\lim S_nx$. \square

The first two equivalences provide an alternative proof of two of the reversals in Theorem 12.5.

16 Final remarks

We hope the explorations here contribute to the growing body of literature on analysis in subsystems of second-order arithmetic, and help show that this is a fertile topic of study. There is much more that can be done; analyses of the spectral theory of Hilbert spaces, as well as properties of more general classes of Banach spaces, would form a natural continuation of the work carried out here.

We have left some loose ends. Most of the open questions have to do with the strength of statements regarding closed linear subsets of a Banach or Hilbert space. Consider the following list:

- (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) For any bounded linear operator from a Banach space to itself and any λ , $\{x \mid Tx = \lambda x\}$ is a closed subspace.
- (6) For any bounded linear operator from a Banach space to itself, $\{x \mid Tx = x\}$ is a closed subspace.
- (7) For any bounded linear operator from a Hilbert space to itself, $\{x \mid Tx = x\}$ is a closed subspace.

In 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 (II_1^I-CA) , and it is also possible that they are all equivalent to (ACA) . It would be nice, therefore, to have a better sense of their logical strength.

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.

Finally, we showed that in RCA_0 each of (WKL) or Σ_2 induction implies that every finite dimensional Banach space has an independent generating sequence. Can one show that this is not provable in RCA_0 outright?

References

- [1] Errett Bishop and Douglas Bridges. *Constructive Mathematics*. Springer-Verlag, Berlin, 1985.
- [2] Douglas Brown. Notions of compactness in weak subsystems of second order arithmetic. To appear in [21].
- [3] Douglas Brown. Notions of closed subsets of a complete separable metric space in weak subsystems of second-order arithmetic. In *Logic and computation*, pages 39–50. American Mathematical Society, Providence, RI, 1990.
- [4] Douglas Brown, Mariagnese Giusto, and Stephen Simpson. Vitali’s theorem and WWKL. *Archive for Mathematical Logic*, 41:191–206, 2002.
- [5] John B. Conway. *A course in functional analysis*. Springer-Verlag, New York, second edition, 1990.
- [6] Harvey Friedman, Stephen Simpson, and Xiaokang Yu. Periodic points and subsystems of second-order arithmetic. *Annals of Pure and Applied Logic*, 62:51–64, 1993.
- [7] Mariagnese Giusto and Alberto Marcone. Lebesgue numbers and Atsugi spaces in subsystems of second-order arithmetic. *Archive for Mathematical Logic*, 37:343–362, 1998.
- [8] Mariagnese Giusto and Stephen Simpson. Located sets and reverse mathematics. *Journal of Symbolic Logic*, 65:1451–1480, 2000.
- [9] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Springer-Verlag, Berlin, 1993.
- [10] David Hilbert and Paul Bernays. *Grundlagen der Mathematik*. Springer-Verlag, Berlin, first volume, 1934, second volume, 1939.
- [11] Jeffry Hirst. Reverse mathematics of separably closed sets. preprint.

- [12] A. James Humphreys and Stephen Simpson. Separation and weak König's lemma. *Journal of Symbolic Logic*, 64:268–278, 1999.
- [13] Alexander Kechris. *Classical descriptive set theory*. Springer-Verlag, New York, 1995.
- [14] Ulrich Kohlenbach. Higher order reverse mathematics. To appear in [21].
- [15] Ulrich Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematics Society*, 357:89–128, 2005.
- [16] Ulrich Kohlenbach and Paulo Oliva. Proof mining in L_1 -approximation. *Annals of Pure and Applied Logic*, 121:1–38, 2003.
- [17] Erwin Kreyszig. *Introductory functional analysis with applications*. John Wiley and Sons, New York, 1978.
- [18] Alberto Marcone. On the logical strength of Nash-Williams' theorem on transfinite sequences. In Wilfrid Hodges, Martin Hyland, Charles Steinhorn, and John Truss, editors, *Logic: from foundations to applications*, pages 327–351. Oxford, Clarendon Press, New York, 1996.
- [19] Karl Petersen. *Ergodic Theory*. Cambridge University Press, Cambridge, 1983.
- [20] Marian Pour-El and J. Ian Richards. *Computability in analysis and physics*. Springer-Verlag, Berlin, 1989.
- [21] Stephen Simpson, editor. *Reverse Mathematics 2001*. Lecture Notes in Logic. ASL Publications, to appear.
- [22] Stephen Simpson. *Subsystems of Second-Order Arithmetic*. Springer-Verlag, Berlin, 1999.
- [23] Bas Spitters. *Constructive and intuitionistic integration theory and functional analysis*. PhD thesis, University of Nijmegen, 2002.
- [24] Gaisi Takeuti. *Two Applications of Logic to Mathematics*. Iwanami Shoten and Princeton University Press, 1978.
- [25] Peter Walters. *An introduction to ergodic theory*. Springer-Verlag, New York, 2000.
- [26] Hermann Weyl. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Veit, Leipzig, 1918. Second edition (1932). Translated by Stephen Pollard and Thomas Bole as *The Continuum. A critical examination of the foundation of analysis*, Dover, New York, 1994.