FUNDAMENTA
MATHEMATICAE
226 (2014)

Normal numbers and the Borel hierarchy

by

Verónica Becher (Buenos Aires), **Pablo Ariel Heiber** (Buenos Aires)
and **Theodore A. Slaman** (Berkeley, CA)

Abstract. We show that the set of absolutely normal numbers is $\mathbf{\Pi}^0_3$-complete in the Borel hierarchy of subsets of real numbers. Similarly, the set of absolutely normal numbers is Π^0_3-complete in the effective Borel hierarchy.

1. Introduction. What is the descriptive complexity of the set of absolutely normal numbers? Alexander Kechris posed this question in the early 1990s when he asked whether the set of real numbers which are normal to base two is $\mathbf{\Pi}^0_3$-complete in the Borel hierarchy. Ki and Linton [5] proved that, indeed, the set of numbers that are normal to any fixed base is $\mathbf{\Pi}^0_3$-complete. However, their proof technique does not extend to the case of absolute normality, that is, normality to all bases simultaneously.

We show that the set of absolutely normal numbers is also $\mathbf{\Pi}^0_3$-complete. In fact, it is Π^0_3-complete in the effective Borel hierarchy. We give, explicitly, a reduction that proves the two completeness results. By a feature of this reduction we also provide an alternate proof of Ki and Linton's theorem. Our analysis here is a refinement of our algorithm for computing absolutely normal numbers in [1].

2. Preliminaries

Notation. As usual, $\mathbb{N}$ is the set of positive integers. A *base* is an integer b greater than or equal to 2, a *digit* in base b is an element of $\{0,\dots,b-1\}$, and a *block* in base b is a finite sequence of digits in base b. The length of a block x is $|x|$. We denote the set of blocks in base b of length ℓ by $\{0,\dots,b-1\}^\ell$. We write the concatenation of two blocks x and u as xu. For arbitrarily many blocks u_i, for $i=1,\dots,m$, $\prod_{1\le i\le m} u_i$ is their concatena-

2010 *Mathematics Subject Classification*: 11U99, 03E15, 68W99.
Key words and phrases: normal numbers, Borel hierarchy, descriptive set theory.

DOI: 10.4064/fm226-1-4

tion in increasing order of i. When the starting value for the index is 1, we abbreviate this expression as $\prod_{i\le m} u_i$.

In case x is a finite or an infinite sequence of digits, $x{\upharpoonright}i$ is the subblock of the first i digits of x, and $x[i]$ is the ith digit of x. A digit d *occurs* in x at position i if $x[i] = d$. A block u *occurs* in x at position i if $x[i+j-1] = u[j]$ for $j = 1, \ldots, |u|$. The number of occurrences of the block u in the block x is $\mathrm{occ}(x, u) = \#\{i : u \text{ occurs in } x \text{ at position } i\}$.

For each real number R in the unit interval we consider its unique expansion in base b, $R = \sum_{i=1}^{\infty} a_i b^{-i}$, where the a_i are digits in base b, and $a_i < b-1$ infinitely many times. This last condition over a_i ensures a unique representation for every rational number. We write $(R)_b$ to denote the expansion of a real R in base b given by the sequence $(a_i)_{i\ge 1}$.

2.1. On normal numbers. Among the several equivalent definitions of absolute normality the following is the most convenient for our presentation. For a reference see the books [2] or [6].

DEFINITION 2.1.

(1) A real number R is *simply normal to base* b if, for each digit d in base b, $\lim_{n\to\infty} \mathrm{occ}((R)_b{\upharpoonright}n, d)/n = 1/b$.
(2) R is *normal to base* b if it is simply normal to the bases b^ℓ for every integer $\ell \ge 1$.
(3) R is *absolutely normal* if it is normal to every base.
(4) R is *absolutely abnormal* if it is non-normal to any base.

Notice that absolute normality is equivalent to being simply normal to every base, but absolute abnormality is not equivalent to being non-simply normal to every integer base.

LEMMA 2.2. *If a real number R is simply normal to at most finitely many bases then R is absolutely abnormal. If a real number R is normal to at most finitely many bases then R is absolutely abnormal.*

Proof. Fix a base b. By contraposition, if R were normal to base b, then it would be simply normal to all bases b^ℓ, a contradiction to the hypothesis of the first claim. Simple normality to all bases b^ℓ implies by definition normality to all bases b^ℓ, which contradicts the hypothesis of the second claim. ∎

The *simple discrepancy* of a block in base b indicates the difference between the actual number of occurrences of the digits in that block and their expected average. The definition of normality can be given in terms of discrepancy (see [2]).

Definition 2.3. Let u be a block of digits in base b. The *simple discrepancy* of the block u to base b is

$$D(u,b) = \max\{|\mathrm{occ}(u,d)/|u| - 1/b| : d \in \{0,\dots,b-1\}\}.$$

Let ℓ be a positive integer. The *block discrepancy* of the block u to blocks of length ℓ in base b is

$$D_\ell(u,b) = \max\{|\mathrm{occ}(u,v)/|u| - 1/b^\ell| : v \in \{0,\dots,b-1\}^\ell\}.$$

Notice that $D(u,b)$ is a number between 0 and $1-1/b$, and $D_\ell(u,b)$ is a number between 0 and $1-1/b^\ell$. By the definition of simple discrepancy, a real number R is simply normal to base b if and only if $\lim_{n\to\infty} D((R)_b{\restriction}n,b) = 0$. Instead of asking for simple discrepancy for every base b^ℓ, $\ell \ge 1$, it is possible to characterize normality using block discrepancy.

Lemma 2.4 ([2, Theorem 4.2]). *A real number R is normal to base b if and only if, for every $\ell \ge 1$, $\lim_{n\to\infty} D_\ell((R)_b{\restriction}n,b) = 0$.*

The next lemma bounds the discrepancy of a concatenation of blocks. We use it very often in what follows, without making explicit reference to it.

Lemma 2.5. *If $u_1,\dots,u_n$ are blocks of digits in base b, then*

$$D\Big(\prod_{j\le n} u_j, b\Big) \le \sum_{j=1}^{n} D(u_j,b)|u_j| \Big/ \sum_{h=1}^{n} |u_h|.$$

Proof. Let d be a digit in base b with $|\mathrm{occ}(\prod_{j\le n} u_j, d)/|\prod_{j\le n} u_j| - 1/b\,|$ maximal. Then

$$\begin{aligned} D\Big(\prod_{j\le n} u_j, b\Big) &= \left|\mathrm{occ}\Big(\prod_{j\le n} u_j, d\Big)\Big/\Big|\prod_{j\le n} u_j\Big| - \frac{1}{b}\right| \\ &\le \left|\Big(\sum_{j=1}^{n} \mathrm{occ}(u_j,d)\Big/\sum_{h=1}^{n}|u_h|\Big) - \frac{1}{b}\right| \\ &\le \left|\sum_{j=1}^{n} \mathrm{occ}(u_j,d) - \frac{|u_j|}{b}\right|\Big/\sum_{h=1}^{n}|u_h| \\ &\le \sum_{j=1}^{n} |u_j|\left|\frac{\mathrm{occ}(u_j,d)}{|u_j|} - \frac{1}{b}\right|\Big/\sum_{h=1}^{n}|u_h| \\ &\le \sum_{j=1}^{n} D(u_j,b)|u_j|\Big/\sum_{h=1}^{n}|u_h|. \ \blacksquare \end{aligned}$$

Borel's fundamental theorem showing that almost all real numbers are absolutely normal is underpinned by the fact that, for any base, almost every sufficiently long block has small simple discrepancy relative to that base. The next lemmas provide an explicit bound for the number of blocks of a given length having simple discrepancy larger than a given value.

Lemma 2.6 ([1, Lemma 2.5] adapted from [3, Theorem 148]). *Let $p_b(k,i)$ be the number of blocks of length k in base b where a given digit occurs exactly i times. Fix a base b and a block length k. For every real ε such that $6/k \le \varepsilon \le 1/b$, $\sum_{i=0}^{k/b-\varepsilon k} p_b(k,i)$ and $\sum_{i=k/b+\varepsilon k}^{k} p_b(k,i)$ are at most $b^k e^{-b\varepsilon^2 k/6}$.*

Lemma 2.7 ([1, Lemma 2.6]). *Let $t \ge 2$ be an integer and let ε and δ be real numbers between 0 and 1, with $\varepsilon \le 1/t$. Let k be the least integer greater than the maximum of $\lceil 6/\varepsilon \rceil$ and $-\ln(\delta/2t)6/\varepsilon^2$. Then, for all $b \le t$ and for all $k' \ge k$, the fraction of blocks x of length k' in base b for which $D(x,b) > \varepsilon$ is less than δ.*

2.2. On descriptive set theory. Recall that the *Borel hierarchy* for subsets of the real numbers is the stratification of the σ-algebra generated by the open sets with the usual interval topology. For references see Kechris's textbook [4] or Marker's lecture notes [7].

A set A is $\boldsymbol{\Sigma}^0_1$ if and only if A is open, and A is $\boldsymbol{\Pi}^0_1$ if and only if A is closed. A is $\boldsymbol{\Sigma}^0_{n+1}$ if and only if it is a countable union of $\boldsymbol{\Pi}^0_n$ sets, and A is $\boldsymbol{\Pi}^0_{n+1}$ if and only if it is a countable intersection of $\boldsymbol{\Sigma}^0_n$ sets.

A is *hard* for a Borel class if and only if every set in the class is reducible to A by a continuous map. A is *complete* in a class if it is hard for this class and belongs to the class.

When we restrict to intervals with rational endpoints and computable countable unions and intersections, we obtain the effective, or lightface, Borel hierarchy. One way to present the finite levels of the effective Borel hierarchy is by means of the arithmetical hierarchy of formulas in the language of second-order arithmetic. Atomic formulas in this language assert algebraic identities between integers or membership of real numbers in intervals with rational endpoints. A formula in the arithmetic hierarchy involves only quantification over integers. A formula is Π^0_0 and Σ^0_0 if all its quantifiers are bounded. It is Σ^0_{n+1} if it has the form $\exists x\, \theta$ where θ is Π^0_n, and it is Π^0_{n+1} if it has the form $\forall x\, \theta$ where θ is Σ^0_n.

A set A of real numbers is Σ^0_n (respectively Π^0_n) in the effective Borel hierarchy if and only if membership in that set is definable by a formula which is Σ^0_n (respectively Π^0_n). Notice that every Σ^0_n set is $\boldsymbol{\Sigma}^0_n$ and every Π^0_n set is $\boldsymbol{\Pi}^0_n$. In fact for every set A in $\boldsymbol{\Sigma}^0_n$ there is a Σ^0_n formula and a real parameter such that membership in A is defined by that Σ^0_n formula relative to that real parameter.

A is *hard* for an effective Borel class if and only if every set in the class is reducible to A by a computable map. As before, A is *complete* in an effective class if it is hard for this class and belongs to the class. Since computable maps are continuous, proofs of hardness in the effective hierarchy often yield proofs of hardness in general by relativization. This is the case in our work.

3. Main theorem. By the form of its definition, normality to a fixed base is explicitly a Π^0_3 property of real numbers. The same holds for absolute normality. Absolute abnormality is, for all bases, the negation of normality, hence a Π^0_4 property.

LEMMA 3.1 (as in [7]). *The set of real numbers that are normal to a given base is* Π^0_3.

The set of real numbers that are absolutely normal is Π^0_3.

The set of real numbers that are absolutely abnormal is Π^0_4.

Thus, to prove completeness of the set of absolutely normal real numbers for the class Π^0_3 we need only prove hardness. We prove our hardness result for the Borel hierarchy by relativizing a hardness result for Π^0_3 subsets of the natural numbers. Let $\mathcal{L}$ be the language of first-order arithmetic. As usual, a *sentence* is a formula without free variables.

THEOREM 3.2. *There is a computable reduction from* Π^0_3 *sentences* φ *in* $\mathcal{L}$ *to indices* e *such that the following implications hold:*

- *If* φ *is true then* e *is the index of a computable absolutely normal number.*
- *If* φ *is false then* e *is the index of a computable absolutely abnormal number.*

We postpone the proofs of Theorem 3.2 and its corollaries to the end of the next section.

COROLLARY 3.3. *The set of absolutely normal numbers is* Π^0_3*-complete, and hence* $\boldsymbol{\Pi}^0_3$*-complete.*

The reduction in Theorem 3.2 gives just two possibilities: absolute normality or absolute abnormality; that is, normality to all bases simultaneously, or to no base at all. Consequently, it also separates normality in base b from non-normality in base b, for any given b.

COROLLARY 3.4. *For every base* b*, the set of normal numbers in base* b *is* Π^0_3*-complete, and hence* $\boldsymbol{\Pi}^0_3$*-complete.*

This gives an alternate proof of Ki and Linton's theorem in [5] for $\boldsymbol{\Pi}^0_3$-completeness, that also covers the case of Π^0_3-completeness. Another consequence of Theorem 3.2 is that absolute abnormality, which is a Π^0_4 property, is hard for the classes Σ^0_3 and $\boldsymbol{\Sigma}^0_3$.

4. Proofs. We shall define a computable reduction that maps Π^0_3 sentences in the language of first order arithmetic $\mathcal{L}$ to indices of computable infinite sequences of zeros and ones. If the given sentence is true then the corresponding binary sequence is the expansion in base 2 of an absolutely normal number. Otherwise, the corresponding binary sequence is the expansion in base 2 of an absolutely abnormal number.

Our reduction is the composition of two reductions. We use Baire space $\mathbb{N}^{\mathbb{N}}$, the set of infinite sequences of positive integers, as an intermediate working space. The first reduction maps sentences from $\mathcal{L}$ to programs that produce infinite sequences of positive integers that reflect the truth or falsity of the given sentences. The second reduction maps these programs to ones that produce binary sequences with the appropriate condition on normality.

By relativizing this reduction, given a Π^0_3 formula in second order arithmetic and a real number X, we produce a binary sequence computably from X which is absolutely normal or absolutely abnormal depending on whether the given formula is true of X. This is exactly what is required to establish Π^0_3-hardness.

4.1. The first reduction. Recall that a Π^0_3 formula in first order arithmetic is equivalent to one starting with a universal quantifier $\forall$, followed by the quantifier "there are only finitely many" $\exists^{<\infty}$ and ended by a computable predicate; see Theorem XVII and Exercise 14-27 in [8]. The computable predicate in this equivalent form comes from the Σ^0_0 subformula of the original.

Definition 4.1. The *first reduction* takes sentences in $\mathcal{L}$ to positive integers by mapping $\forall x \exists^{<\infty} y\ C(x, y)$, where C is computable, to an index of the following program: For every positive integer n in increasing order, let $x = \max\{z \in \mathbb{N} : 2^z \text{ divides } n\}$ and let $y = n/2^x$. If $y = 1$ or $C(x, y)$ then append $x, x+1, \ldots, x+y-1$ to the output sequence.

Lemma 4.2. *If φ is a Π^0_3 sentence in $\mathcal{L}$ then Definition* 4.1 *gives the index of a program that outputs an infinite sequence f of integers such that the subsequence of f's first occurrences is an enumeration of $\mathbb{N}$ in increasing order and the following dichotomy holds:*

- *If φ is true then no positive integer occurs infinitely often in f.*
- *If φ is false then all but finitely many integers occur infinitely often in f.*

Proof. Assume φ of the form $\forall x \exists^{<\infty} y\ C(x, y)$, where C is computable. We say that a tuple $\langle x, y \rangle$ is *appending* if $y = 1$ or $C(x, y)$. It is clear by inspection that all possible pairs $\langle x, y \rangle$ with $x, y \in \mathbb{N}$ are processed, and that $\langle x+1, y \rangle$ and $\langle x, y+1 \rangle$ are always processed after $\langle x, y \rangle$.

Thus, the first occurrence of an integer $x+1$ in f is due to the appending tuple $\langle x+1, 1\rangle$ or an appending tuple $\langle x', y\rangle$ with $x' < x+1$ and $x'+y > x+1$. In the first case, the appending tuple $\langle x+1, 1\rangle$ is processed after $\langle x, 1\rangle$, which appends x to the output. In the second case, the processing of an appending tuple $\langle x', y\rangle$ with $x' < x+1$ and $x'+y > x+1$ appends x right before $x+1$ to the output. Thus, $x+1$ occurs for the first time in f after x.

Suppose now φ is true. For any x, there are finitely many appending tuples of the form $\langle x', y\rangle$ with $x' \leq x$. After all such appending tuples have been processed, x will not be appended to the output sequence. Thus no positive integer can occur infinitely often in f.

Now suppose φ is false. Let x be such that there are infinitely many y such that $C(x, y)$. Let z be any positive integer. Each time an appending tuple of the form $\langle x, y\rangle$ with $z < y$ is processed, $x + z$ is appended to the output. Since we assumed there are infinitely many such tuples, $x + z$ is appended to the output an infinite number of times. Thus, all integers greater than x occur infinitely often in the output sequence. ∎

4.2. The second reduction. We use the phrase *b-adic interval* to refer to a semi-open interval of the form $[a/b^m, (a+1)/b^m)$ for $a < b^m$. We move freely between b-adic intervals and base-b expansions. If x is a base-b block and it is understood that we are working in base b, then we let $.x$ denote the rational number whose expansion in base b has exactly the digits occurring in x. Given the block x, the reals such that their base-b expansions extend x are exactly those belonging to the b-adic interval $[.x, .x + b^{-|x|})$. Conversely, every b-adic interval $[a/b^m, (a+1)/b^m)$ corresponds to a block x as above, where x is obtained by writing a in base b and then prepending a sufficient number of zeros to obtain a block of length m. We use μ to denote Lebesgue measure and log to denote logarithm in base 2.

Now we introduce some tools and establish a few of their properties.

LEMMA 4.3 ([1]). *For every non-empty interval I and base b, there is a b-adic subinterval I_b of I such that $\mu(I_b) \geq \mu(I)/(2b)$. Moreover, such a subinterval can be computed uniformly from I and b.*

DEFINITION 4.4 ([1]). A *t-sequence* is a nested sequence of $t-1$ intervals, $\vec{I} = (I_2, \ldots, I_t)$, such that I_2 is dyadic and for each base $b \leq t$, I_{b+1} is a $(b+1)$-adic subinterval of I_b such that $\mu(I_{b+1}) \geq \mu(I_b)/2(b+1)$. We let $x_b(\vec{I})$ be the block in base b such that $.x_b(\vec{I})$ is the expansion of the left endpoint of I_b in base b.

We can iteratively apply Lemma 4.3 to obtain the following corollary.

COROLLARY 4.5. *For every non-empty dyadic interval I and every integer $t \geq 2$, there is a t-sequence that begins with I. Moreover, such a t-sequence can be computed uniformly from I and t.*

If $\vec{I} = (I_2, \ldots, I_t)$ is a t-sequence, then for any base $b \leq t$ and any real $X \in I_t$, X has $x_b(\vec{I})$ as an initial segment of its expansion in base b. If, further, $\vec{I'} = (I'_2, \ldots, I'_{t'})$ is a t'-sequence with $t \leq t'$ such that $I'_2 \subset I_t$ and $X \in I'_{t'}$, then for each $b \leq t$, $\vec{I'}$ specifies how to extend $x_b(\vec{I})$ to a longer initial segment $x_b(\vec{I'})$ of the base-b expansion of X. As opposed to arbitrary nested sequences, for t-sequences there is a function that gives a lower bound for the ratio between the measures of I_b and $I_{b'}$, for any two bases b and b' both less than or equal to t. That is, assuming $b > b'$, we have

$$\mu(I_b) \geq \frac{\mu(I_{b'})}{2^{b-b'} b!/b'!}.$$

In what follows we use the above inequality repeatedly.

Our next task is to define a function that, for an integer i and a t-sequence $\vec{I}$ (for some t), constructs an $(i+1)$-sequence inside $\vec{I}$ with good properties. The method is to determine the expansion of the rational endpoints of each b-adic interval in the $(i+1)$-sequence. Since the respective b-adic intervals are nested, the determination of the expansions is done by adding suffixes.

We introduce three functions of i, $(\delta_i)_{i\geq 1}$, $(k_i)_{i\geq 1}$ and $(\ell_i)_{i\geq 1}$, that act as parameters for the construction of an $(i+1)$-sequence. The integer k_i indicates how many digits in base $i+1$ can be determined in each step; thus, $k_i\lceil \log(i+1)\rceil$ indicates how many digits in base 2 can be determined in each step (keep in mind that, in general, more digits are needed to ensure the same precision in base 2 than in a larger base). The integer ℓ_i limits how many digits in base 2 can *at most* be determined in each step. And the rational δ_i bounds the relative measure of any two intervals in two consecutive nested $(i+1)$-sequences. We let $\textsc{Ref}_i$ be the function that given $\vec{I}$ constructs an $(i+1)$-sequence by recursion. It is a search through nested $(i+1)$-sequences until one with good properties is reached. The choices we make for $(\delta_i)_{i\geq 1}$, $(k_i)_{i\geq 1}$ and $(\ell_i)_{i\geq 1}$ allow us to prove the correctness of the construction.

Definition 4.6. Let $(k_i)_{i\geq 1}$ and $(\ell_i)_{i\geq 1}$ be the computable sequences of positive integers and let $(\delta_i)_{i\geq 1}$ be the computable sequence of positive rational numbers less than 1 such that, for each $i \geq 1$,

$$\delta_i = \frac{1}{2^{2i-2}(i+1)!^2},$$

$$k_i = \text{least integer greater than}$$

$$\max\left(\lceil 6(i+2)\rceil, -\ln\left(\frac{\delta_i}{2(i+1)^2}\right)6(i+2)^2\right),$$

$$\ell_i = k_i\lceil \log(i+1)\rceil + \lceil -\log(\delta_i)\rceil.$$

DEFINITION 4.7. REF_i maps a $(p+1)$-sequence $\vec{I} = (I_2, \ldots, I_{p+1})$ into an $(i+1)$-sequence $\text{REF}_i(\vec{I})$ that we define recursively:

Initial step 0. Let $\vec{I_0} = (I_{0,2}, \ldots, I_{0,i+1})$ be an $(i+1)$-sequence where $I_{0,2}$ is the leftmost dyadic subinterval of I_{p+1} such that $\mu(I_{0,2}) \geq \mu(I_{p+1})/4$.

Recursive step $j+1$. Let $\vec{I}_{j+1}$ be the $(i+1)$-sequence such that:

- Let L be the leftmost dyadic subinterval of $I_{j,i+1}$ such that
$$\mu(L) \geq \mu(I_{j,i+1})/4.$$
- Partition L into $k_i\lceil\log(i+1)\rceil$ dyadic subintervals of equal measure $2^{-k_i\lceil\log(i+1)\rceil}\mu(L)$. For each such subinterval J_2 of L, define the $(i+1)$-sequence $\vec{J} = (J_2, \ldots, J_{i+1})$.
- Let $\vec{I}_{j+1}$ be the leftmost of the $(i+1)$-sequences $\vec{J}$ considered above such that $D(u_b(\vec{J}), b) \leq 1/(i+2)$ for each base $b \leq i+1$, where $u_b(\vec{J})$ is such that $x_b(\vec{I_j})u_b(\vec{J}) = x_b(\vec{J})$.

Repeat the recursion until step n when all the following hold for every base $b \leq i+1$:

(a) $|x_b(\vec{I_n})| > \ell_{i+1}(i+3)$,
(b) $D(x_b(\vec{I_n}), b) \leq 2/(i+2)$,
(c) $D_{2\ell_i}(x_b(\vec{I_n}), b) > b^{-2\ell_i-1}$.

Finally, let $\text{REF}_i(\vec{I}) = \vec{I_n}$.

The following lemmas show that for every positive integer i, the function REF_i is well defined and it is computable.

LEMMA 4.8. *There is always a suitable $(i+1)$-sequence $\vec{J}$ to be selected in the recursive step of Definition* 4.7.

Proof. Consider the recursive step $j+1$ of Definition 4.7. Let S be the union of the set of intervals J_{i+1} over the $2^{k_i\lceil\log(i+1)\rceil}$ $(i+1)$-sequences $\vec{J}$. We have $\mu(L) \geq \mu(I_{j,i+1})/4$ and, as $\vec{J}$ and $\vec{I_j}$ are $(i+1)$-sequences,
$$\mu(J_{i+1}) \geq \frac{\mu(J_2)}{2^{i-2}(i+1)!} \quad \text{and} \quad \mu(I_{j,i+1}) \geq \frac{\mu(I_{j,2})}{2^{i-2}(i+1)!}.$$
Since the possibilities for J_2 form a partition of L,
$$\mu(S) \geq \frac{\mu(L)}{2^{i-2}(i+1)!} \geq \frac{\mu(I_{j,i+1})}{2^{i}(i+1)!} \geq \frac{\mu(I_{j,2})}{2^{2i-2}(i+1)!^2} = \delta_i\mu(I_{j,2}).$$
Let us say that an $(i+1)$-sequence $\vec{J}$ is *not suitable* if for some base $b \leq i+1$,
$$D(u_b(\vec{J}), b) > 1/(i+2).$$

Let N be the subset of S defined as the union of the set of intervals J_{i+1} which occur in $(i+1)$-sequences which are not suitable. Each $\vec{J}$ considered at stage $i+1$ is such that for every base $b \le i+1$ each interval J_b is a subinterval of $I_{j,b}$. By definition, $|u_b(\vec{J})| > k_i$ for each b and $\vec{J}$. By Lemma 2.7 with $t = i+1$, $\varepsilon = 1/(i+2)$, $\delta = \delta_i/(i+1)$ and $k = k_i$, for each base $b \le i+1$, the subset of $I_{j,b}$ consisting of all reals with base-b expansions starting with $x_b(\vec{I_j})u_b(\vec{J})$ for which $D(u_b(\vec{J}), b) > 1/(i+2)$ has measure less than $\delta\mu(I_{j,b})$, and hence less than $\delta\mu(I_{j,2})$. Therefore,

$$\mu(N) < (i+1)\delta\mu(I_{j,2}) = \delta_i \mu(I_{j,2}) \le \mu(S).$$

This proves that S is a proper superset of N, therefore, there is a suitable $(i+1)$-sequence. ■

LEMMA 4.9. *The recursion in Definition* 4.7 *finishes for every input sequence* $\vec{I}$.

Proof. Using Lemma 4.3 or Corollary 4.5 the needed b-adic subintervals with the appropriate measure and $(i+1)$-sequences can be found computably. Lemma 4.8 ensures that a suitable $\vec{J}$ can always be found in each recursive step. All the other tasks in the recursive step are clearly computable. It remains to check that the ending conditions of the recursion are eventually met. Let $u_{b,j+1}$ be such that

$$x_b(\vec{I_j})u_{b,j+1} = x_b(\vec{I}_{j+1}),$$

and v_b be such that

$$x_b(\vec{I_0}) = x_b(\vec{I})v_b.$$

Then $1 \le |u_{b,j}|$ and

$$\begin{aligned}
|u_{b,j}| &= |x_b(\vec{I_j})| - |x_b(\vec{I}_{j-1})| = -\log_b \frac{\mu(I_{j,b})}{\mu(I_{j-1,b})} \\
&= -\log_b \left(\frac{\mu(I_{j,b})}{\mu(I_{j,2})} \, \frac{\mu(I_{j,2})}{\mu(I_{j-1,i+1})} \, \frac{\mu(I_{j-1,i+1})}{\mu(I_{j-1,b})} \right) \\
&\le -\log_b \left(\frac{1}{2^{b-3} b!} \, \frac{1}{4 \cdot 2^{k_i \lceil \log(i+1) \rceil}} \, \frac{1}{2^{i+1-b}(i+1)!/b!} \right) \\
&\le -\log \left(\frac{1}{2^{i-2}(i+1)!} \, \frac{1}{4 \cdot 2^{k_i \lceil \log(i+1) \rceil}} \right) \le k_i \lceil \log(i+1) \rceil - \log \delta_i \le \ell_i.
\end{aligned}$$

The recursive step establishes that $D(u_{b,j}, b) \le 1/(i+2)$, and also that for any k, $x_b(\vec{I_k}) = x_b(\vec{I})v_b \prod_{j \le k} u_{b,j}$. Notice that, in each of the three conditions (a), (b) and (c), the right side of the inequality is fixed. For condition (a), $|x_b(\vec{I_n})| = |x_b(\vec{I})v_b \prod_{j \le n} u_{b,j}|$ is strictly increasing in n, so it is greater than the required lower bound for sufficiently large n. For condition (b), observe that

$$D(x_b(\vec{I_n}), b) = D\Big(x_b(\vec{I})v_b \prod_{j\le n} u_{b,j}, b\Big) \le |x_b(\vec{I})v_b|/|x_b(\vec{I_n})| + D\Big(\prod_{j\le n} u_{b,j}, b\Big)$$
$$\le |x_b(\vec{I})v_b|/|x_b(\vec{I_n})| + 1/(i+2).$$

On the right hand side, the first term approaches 0 for large n, so the entire expression is less than $2/(i+2)$ for sufficiently large n. For condition (c), observe that the recursive step ensures that $u_{b,j}$ is never all zeros. So, a sequence of $2\ell_i$ zeros does not occur in $\prod_{j\le n} u_{b,j}$. By definition,

$$D_{2\ell_i}\Big(x_b(\vec{I})v_b \prod_{j\le n} u_{b,j}, b\Big) \ge \left|\frac{\mathrm{occ}(x_b(\vec{I})v_b \prod_{j\le n} u_{b,j}, 0^{2\ell_i})}{|x_b(\vec{I})v_b \prod_{j\le n} u_{b,j}|} - \frac{1}{b^{2\ell_i}}\right|.$$

Since $\mathrm{occ}(x_b(\vec{I})v_b \prod_{j\le n} u_{b,j}, 0^{2\ell_i})$ is bounded by a constant, for sufficiently large n, the discrepancy $D_{2\ell_i}(x_b(\vec{I})v_b \prod_{j\le n} u_{b,j}, b)$ is arbitrarily close to $b^{-2\ell_i}$. ■

LEMMA 4.10. *Let $\vec{I}$ be an arbitrary $(p+1)$-sequence, $i \ge 1$ be an integer and $\vec{R}$ be* $\mathrm{REF}_i(\vec{I})$. *For every base $b \le \min(i,p)+1$,*

(1) $R_2 \subseteq I_{p+1}$,
(2) $D(x_b(\vec{R}), b) \le 2/(i+2)$,
(3) $D_{2\ell_i}(x_b(\vec{R}), b) > b^{-2\ell_i-1}$,
(4) $|x_b(\vec{R})| > \ell_{i+1}(i+3)$,
(5) *for each ℓ such that $|x_b(\vec{I})| \le \ell \le |x_b(\vec{R})|$,*

$$D(x_b(\vec{R})\restriction\ell, b) \le D(x_b(\vec{I}), b) + \lceil -\log(\delta_p)\rceil/|x_b(\vec{I})| + 1/(i+2) + \ell_i/|x_b(\vec{I})|.$$

Proof. As in the proof of Lemma 4.9, let $u_{b,j+1}$ be such that

$$x_b(\vec{I_j})u_{b,j+1} = x_b(\vec{I}_{j+1}),$$

and let v_b be such that

$$x_b(\vec{I_0}) = x_b(\vec{I})v_b.$$

Then $1 \le |u_{b,j}| \le \ell_i$, $D(u_{b,j}, b) \le 1/(i+2)$, and for any k we find that $x_b(\vec{I_k}) = x_b(\vec{I})v_b \prod_{j\le k} u_{b,j}$.

Fix a base b. Point (1) follows by induction on the recursive steps in the definition of $\mathrm{REF}_i(\vec{I})$, since each subsequent interval is contained in the previous one. Points (2), (3) and (4) follow from the termination condition in that recursion. For (5), use the above definition of v_b and the parameter

δ_p (see Definition 4.6):

$$\begin{aligned}|v_b| &= |x_b(\vec{I_0})| - |x_b(\vec{I})| = -\log_b \frac{\mu(I_{0,b})}{\mu(I_b)}\\ &= -\log_b\left(\frac{\mu(I_{0,b})}{\mu(I_{0,2})}\,\frac{\mu(I_{0,2})}{\mu(I_{p+1})}\,\frac{\mu(I_{p+1})}{\mu(I_b)}\right)\\ &= -\log_b\left(\frac{1}{2^{b-3}b!}\,\frac{1}{4}\,\frac{1}{2^{p+1-b}(p+1)!/b!}\right) = -\log\frac{1}{2^p(p+1)!}\\ &\le -\log\delta_p \le \lceil -\log(\delta_p)\rceil.\end{aligned}$$

Then, for each m, $D(\prod_{j\le m} u_{b,j}, b) \le 1/(i+2)$ and

$$D\Big(x_b(\vec{I})v_b\prod_{j\le m} u_{b,j}, b\Big) \le D(x_b(\vec{I}), b) + \lceil -\log(\delta_p)\rceil/|x_b(\vec{I})| + 1/(i+2).$$

Finally, fix ℓ and let m and ℓ' be such that $(x_b(\vec{I})v_b\prod_{j\le m} u_{b,j})(u_{b,m+1}\restriction\ell') = x_b(\vec{R})\restriction\ell$. Then

$$\begin{aligned}D(x_b(\vec{R})\restriction\ell, b) &= D\Big(\Big(x_b(\vec{I})v_b\prod_{j\le m} u_{b,j}\Big)(u_{b,m+1}\restriction\ell'), b\Big)\\ &\le D(x_b(\vec{I}), b) + \lceil -\log(\delta_p)\rceil/|x_b(\vec{I})| + 1/(i+2) + |u_{b,m+1}|/|x_b(\vec{I})|\\ &\le D(x_b(\vec{I}), b) + \lceil -\log(\delta_p)\rceil/|x_b(\vec{I})| + 1/(i+2) + \ell_i/|x_b(\vec{I})|. \blacksquare\end{aligned}$$

Definition 4.11. We define the function LimRef that takes infinite sequences of positive integers to real numbers in the unit interval, and

$$\text{LimRef}(f) \text{ is the unique element in } \bigcap_{j=1}^{\infty}(\vec{R}_j)_2,$$

where $\vec{R}_0 = ([0,1))$ and $\vec{R}_{j+1} = \text{Ref}_{f_{j+1}}(\vec{R}_j)$.

That is, LimRef(f) is the real obtained by iterating applications of Ref$_i$ where i is determined by the positive integers in f. By point (1) of Lemma 4.10, for each $j \ge 1$, LimRef(f) is inside every interval in every (f_j+1)-sequence $\vec{R}_j$, and therefore, for each base $b \le f_j + 1$, $x_b(\vec{R}_j)$ is a prefix of $(\text{LimRef}(f))_b$.

Lemma 4.12. *Let f be a sequence of positive integers such that the subsequence of f's first occurrences is an enumeration of $\mathbb{N}$ in increasing order and no positive integer occurs infinitely often in f. Then* LimRef(f) *is an absolutely normal number.*

Proof. Fix a base b and $\varepsilon > 0$. We prove that $D((\text{LimRef}(f))_b\restriction\ell, b) \le \varepsilon$ for each sufficiently large ℓ. Let j_0 be large enough such that $f_j > \max(b, \lceil 8/\varepsilon\rceil)$ for every $j \ge j_0$. Consider $\ell > |x_b(\vec{R}_{j_0})|$, and noticing that $(|x_b(\vec{R}_j)|)_{j\in\mathbb{N}}$ is an increasing sequence, let j be such that $|x_b(\vec{R}_j)| \le \ell < |x_b(\vec{R}_{j+1})|$. Observe that $(\textsc{LimRef}(f))_b\restriction\ell = x_b(\vec{R}_{j+1})\restriction\ell$. Now note that $1/(f_j+2) \le \varepsilon/8$ and apply point (2) of Lemma 4.10 to $\vec{R}_j = \textsc{Ref}_{f_j}(\vec{R}_{j-1})$ to conclude that

$$D(x_b(\vec{R}_j), b) \le 2/(f_j+2) \le \varepsilon/4.$$

By hypothesis, $f_j, f_{j+1} > b > 1$, so let $j_1 < j$ be such that $f_{j_1} = f_j - 1$ and $j_2 < j+1$ be such that $f_{j_2} = f_{j+1} - 1$. By point (4) of Lemma 4.10,

$$|x_b(\vec{R}_j)| \ge |x_b(\vec{R}_{j_1})| > \ell_{f_j}(f_j+2) > \lceil -\log(\delta_{f_j})\rceil (f_j+2),$$

and so

$$\lceil -\log(\delta_{f_j})\rceil / |x_b(\vec{R}_j)| < 1/(f_j+2) \le \varepsilon/8.$$

Similarly,

$$|x_b(\vec{R}_j)| \ge |x_b(\vec{R}_{j_2})| > \ell_{f_{j+1}}(f_{j+1}+2),$$

hence

$$\ell_{f_{j+1}}/|x_b(\vec{R}_j)| < 1/(f_{j+1}+2) \le \varepsilon/8.$$

Now consider Lemma 4.10 again, but applied to $\vec{R}_{j+1} = \textsc{Ref}_{f_{j+1}}(\vec{R}_j)$. By point (5),

$$\begin{aligned} D(x_b(\vec{R}_{j+1})\restriction\ell, b) \le D(x_b(\vec{R}_j), b) &+ \lceil -\log(\delta_{f_j})\rceil/|x_b(\vec{R}_j)| \\ &+ 1/(f_j+2) + \ell_{f_{j+1}}/|x_b(\vec{R}_j)|. \end{aligned}$$

By the bounds established above, each term on the right side of the inequality is at most $\varepsilon/4$. So,

$$D((\textsc{LimRef}(f))_b\restriction\ell, b) = D(x_b(\vec{R}_{j+1})\restriction\ell, b) \le \varepsilon.$$

By the choice of j_0, j, j_1, j_2 the sequences $\vec{R}_{j_0}, \vec{R}_j, \vec{R}_{j_1}, \vec{R}_{j_2}$ contain a b-adic interval, hence the function x_b is defined on them. ∎

LEMMA 4.13. *Let f be a sequence of positive integers such that all but finitely many occur infinitely often in f. Then* LIMREF(f) *is absolutely abnormal.*

Proof. Fix a base b such that b appears infinitely often in f. By the conditions imposed on f, $(\textsc{LimRef}(f))_{b+1}$ has infinitely many prefixes of the form $x_{b+1}(\textsc{Ref}_b(\vec{I}))$ for some $\vec{I}$. By point (3) of Lemma 4.10,

$$D_{2\ell_b}(x_{b+1}(\textsc{Ref}_b(\vec{I})), b+1) > (b+1)^{-2\ell_b-1}.$$

Hence, for infinitely many prefixes of $(\textsc{LimRef}(f))_{b+1}$ their discrepancy to blocks of length $2\ell_b$ in base $b+1$ is bounded away from 0. Then, by Lemma 2.4, LIMREF(f) is not normal to base $b+1$. Since all but finitely many bases can be chosen as $b+1$, LIMREF(f) is not normal to all but finitely many bases. By Lemma 2.2 it is absolutely abnormal. ∎

We are ready to define the second reduction.

Definition 4.14. The *second reduction* maps the index for a computable infinite sequence of integers f to the index for the infinite binary sequence $(\textsc{LimRef}(f))_2$.

Since $\textsc{LimRef}(f)$ is uniformly computable from the input f, the second reduction is computable.

4.3. Proof of the main theorem and corollaries

Proof of Theorem 3.2. The needed reduction is the composition of the first reduction, given in Definition 4.1, and the second reduction, given in Definition 4.14. Apply Lemma 4.2 for the first reduction and Lemma 4.12 for the second reduction to obtain the first implication in Theorem 3.2. Apply Lemma 4.2 for the first reduction and Lemma 4.13 for the second reduction to obtain the second implication. ∎

Proof of Corollary 3.3. Lemma 3.1 states that the corresponding sets are in the Π^0_3 and $\boldsymbol{\Pi}^0_3$ classes. The hardness result in the effective case is immediate from Theorem 3.2 by relativization. We have the reduction from a Π^0_3 sentence in first order arithmetic to an appropriate index for a computable real number. By relativization, we obtain a reduction from a Π^0_3 statement about a real number X to an appropriate index of a real number which is computable from X.

For the general case, recall that to prove hardness of subsets of reals at levels in the Borel hierarchy it is sufficient to consider subsets of Baire space $\mathbb{N}^\mathbb{N}$, because there is a continuous function from the real numbers to $\mathbb{N}^\mathbb{N}$ that preserves $\boldsymbol{\Pi}^0_3$ definability. Baire space admits a syntactic representation of the levels in the Borel hierarchy in arithmetical terms, namely a subset of $\mathbb{N}^\mathbb{N}$ can be defined by a Π^0_3 formula with a fixed parameter $P \in \mathbb{N}^\mathbb{N}$. The analysis given for the effective case, but now relativized to X and P, applies. ∎

Proof of Corollary 3.4. Observe that normality in all bases implies normality in each base. And absolute abnormality is lack of normality in every base. Thus, the same reductions used in the proof of Corollary 3.3 also prove the completeness results for just one fixed base. ∎

Acknowledgements. This research received support from CONICET and Agencia Nacional de Promoción Científica y Tecnológica, Argentina, from the National Science Foundation, USA, under Grant No. DMS-1001551, and from the Simons Foundation. V. Becher and P. A. Heiber are members of the Laboratoire International Associé INFINIS, CONICET/Universidad de Buenos Aires–CNRS/Université Paris Diderot. This work was done while the authors participated in the Buenos Aires Semester in Computability, Complexity and Randomness, 2013.

References

[1] V. Becher, P. A. Heiber, and T. A. Slaman, *A polynomial-time algorithm for computing absolutely normal numbers*, Inform. and Comput. 232 (2013), 1–9.

[2] Y. Bugeaud, *Distribution Modulo One and Diophantine Approximation*, Cambridge Tracts in Math. 193, Cambridge Univ. Press, Cambridge, 2012.

[3] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th ed., Oxford Univ. Press, Oxford, 2008.

[4] A. S. Kechris, *Classical Descriptive Set Theory*, Grad. Texts in Math. 156, Springer, New York, 1995.

[5] H. Ki and T. Linton, *Normal numbers and subsets of* $\mathbb{N}$ *with given densities*, Fund. Math. 144 (1994), 163–179.

[6] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Dover, Mineola, NY, 2006.

[7] D. Marker, *Descriptive set theory*, course notes, 2002.

[8] H. Rogers, Jr., *Theory of Recursive Functions and Effective Computability*, 2nd ed., MIT Press, Cambridge, MA, 1987.

Verónica Becher
Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
and
CONICET
Pabellón I, Ciudad Universitaria
1428 Buenos Aires, Argentina
E-mail: vbecher@dc.uba.ar

Pablo Ariel Heiber
Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Pabellón I, Ciudad Universitaria
1428 Buenos Aires, Argentina
E-mail: pheiber@dc.uba.ar

Theodore A. Slaman
Department of Mathematics
The University of California, Berkeley
719 Evans Hall #3840
Berkeley, CA 94720-3840, U.S.A.
E-mail: slaman@math.berkeley.edu

Received 20 June 2013;
in revised form 16 January 2014