On the Vector Space of the Automatic Reals

(Extended Abstract)

Siegfried Lehr[*], Jeffrey Shallit[†], and John Tromp[‡]

Abstract

A sequence $(a_n)_{n\geq 0}$ is said to be k-automatic if a_n is a finite-state function of the base-k digits of n. We say a real number is (k,b)-automatic if its fractional part has a base-b expansion that forms a k-automatic sequence, and we denote the set of all such numbers as $L(k,b)$. Lehr [*Theoret. Comput. Sci.* **108** (1993), 385–391] proved that $L(k,b)$ forms a vector space over $\mathbb{Q}$. In this paper we give a shortened version of the proof of Lehr's result and, answering a question of Bach, show that the dimension of the vector space $L(k,b)$ is infinite.

We also give an example of a transcendental number such that all of its positive powers are automatic. The proof requires examining the coefficient of X^n in the formal power series $(X+X^2+X^4+X^8+\cdots)^r$. Along the way we are led to examine several sequences of independent combinatorial interest.

Finally, we solve the open problem of whether or not the automatic reals are closed under product by exhibiting a counterexample.

Résumé

On appelle une suite $(a_n)_{n\geq 0}$ k-automatique si a_n est une fonction d'état fini des chiffres de n en base k. On appelle un nombre réel y (k,b)-automatique si le développement en base b de la partie fractionnaire de y est une suite k-automatique, et on écrit $L(k,b)$ pour l'ensemble de ces nombres y. Lehr [*Theoret. Comput. Sci.* **108** (1993), 385–391] a prouvé que $L(k,b)$ est un espace vectoriel sur $\mathbb{Q}$. Dans cet article, nous donnons une version abrégée de la démonstration du résultat de Lehr, et nous montrons que la dimension de l'espace vectoriel $L(k,b)$ est infini (résolution d'une question de Bach).

Nous donnons aussi l'exemple d'un nombre réel transcendant y tel que $y, y^2, y^3, \ldots$ sont tous des nombres automatiques. La démonstration nécessite le calcul du coefficient de X^n dans la série formelle $(X + X^2 + X^4 + X^8 + \cdots)^r$. Nous étudions aussi quelques suites apparentées, qui ont un intérêt combinatoire.

Enfin, nous démontrons que l'ensemble des nombres réels automatiques n'est pas fermé pour le produit (résolution d'une question ouverte).

1 Introduction.

Let $(a_n)_{n\geq 0}$ be an infinite sequence over a finite alphabet Δ. Then we say $(a_n)_{n\geq 0}$ is k-automatic if, roughly speaking, a_n is a finite-state function of the base-k expansion of n.

[*]Mathematisches Institut A, Universität Stuttgart, Pfaffenwaldring 57, D-70511 Stuttgart, Germany. E-mail: lehr@fermat.mathematik.uni-stuttgart.de.

[†]Department of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail: shallit@graceland.uwaterloo.ca. Supported in part by a grant from NSERC Canada. All correspondence should be directed to this author.

[‡]Department of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail: tromp@math.uwaterloo.ca. Supported in part by an NSERC International Fellowship and ITRC.

More precisely, we define a *deterministic finite automaton with output* (DFAO) to be a 6-tuple $M = (Q, \Sigma, \delta, q_0, \Delta, \tau)$, where Q is a finite set of states, Σ is the finite input alphabet, $\delta : Q \times \Sigma \to Q$ is the transition function, q_0 is the initial state, Δ is the finite output alphabet, and $\tau : Q \to \Delta$ is the output function. On input w, the *output* of M is defined to be $\tau(\delta(q_0, w))$. (For more information on automata theory, see [9].)

Then our formal definition of a *k-automatic sequence* is as follows: the sequence $(a_n)_{n\geq 0}$ is k-automatic if there exists a DFAO such that, for all integers $n \geq 0$, we have $\tau(\delta(q_0, (n)_k)) = a_n$. If this is the case, then we say that the DFAO generates $(a_n)_{n\geq 0}$. Here $(n)_k$ is defined to be the *standard base-k representation of n* over the alphabet $\Sigma = \{0, 1, \ldots, k-1\}$, written with the most significant digit at the left, and with no leading zeroes. Note that $(0)_k = \epsilon$, the empty string.

Now let y be a real number, b be an integer ≥ 2, and suppose

$$y \equiv \sum_{i\geq 0} a_i b^{-i-1} \pmod 1,$$

where the a_i are integers satisfying $0 \leq a_i < b$. That is, the sequence $(a_i)_{i\geq 0}$ gives the base-b expansion of $\{y\}$, the fractional part of y. (Technically speaking, we also allow the case where $a_i = b-1$ for all $i \geq 0$.) If $(a_i)_{i\geq 0}$ is a k-automatic sequence, then we say that y is a (k, b)-automatic real number. The set of all such numbers is denoted by $L(k, b)$.

Lehr [12] proved that $L(k, b)$ is a vector space over the rationals, but his proof was somewhat more complicated than necessary. In Section 2, we simplify Lehr's proof, and generalize it somewhat. In Section 3 we give a simple proof that $L(k, b)$ is of infinite dimension over $\mathbb{Q}$. In Section 4 we consider the question of producing a single transcendental number y such that $\mathbb{Q}[y] \subseteq L(k, b)$; our proof requires examining $\alpha(r, n)$, the coefficient of X^n in the formal power series $(X + X^2 + X^4 + X^8 + \cdots)^r$. In Section 5 we give one proof that $\alpha(r, n)$ is bounded for each fixed r. The proof is based on a relationship to a previously-studied sequence that we explore in more detail in Section 6. In Section 7, we provide an improved bound on $\alpha(r, n)$, and exhibit a relationship with the Catalan tree of J. West. Finally, in Section 8, we show that $L(k, b)$ is not closed under product; hence, it is not a ring.

2 $L(k, b)$ is a vector space over $\mathbb{Q}$.

In this section, we reprove Lehr's result that $L(k, b)$ is a vector space over the rationals. To do this, it suffices to show that if y, y' are in $L(k, b)$, and n is any positive integer, then each of (i) $-y$; (ii) y/n; and (iii) $y + y'$ are also in $L(k, b)$. The first is easy. The second follows immediately from the observation that long division by n is a uniform finite transduction, and Cobham [3, Thm. 4] proved that automatic sequences are closed under this type of transduction. The third statement will follow from the following slightly more general lemma.

Lemma 1 (The Normalization Lemma) *Let $(a_i)_{i\geq 0}$ be a bounded k-automatic sequence of non-negative integers. Let $C = \sup_{i\geq 0} a_i$. Then $y = \sum_{i\geq 0} a_i b^{-i-1}$ is a (k, b)-automatic real number.*

It is perhaps worth emphasizing that this result is *not* related to the normalization results of Frougny (see, e.g., [7]).

Proof. The result is trivial if $C < b$, for then the digits in the base-b expansion of y are precisely a_i. The only difficulty occurs when the carries are taken into account, since carries may come from arbitrarily far to the right.

The idea of the proof is as follows: first, in a bounded number of steps, we rewrite $y = \sum_{i \geq 0} a'_i b^{-i-1}$ in such a way that $0 \leq a'_i \leq b$. Next, we show how perform the potential carries resulting from the digits equal to b.

For the first step, define $g_i = a_i \bmod b$, and $h_i = \lfloor a_{i+1}/b \rfloor$ for $i \geq 0$. Then clearly $y \equiv \sum_{i \geq 0} a'_i b^{-i-1} \pmod 1$, where $a'_i = g_i + h_i$. Now $(g_i)_{i \geq 0}$ is easily seen to be k-automatic, and the fact that $(h_i)_{i \geq 0}$ is k-automatic follows from a remark of Cobham [3, p. 174]. Hence $(a'_i)_{i \geq 0}$ is k-automatic.

Now if $a_i \leq C$ for all $i \geq 0$, then $a'_i \leq b - 1 + \lfloor C/b \rfloor$. By repeating this transformation at most $\lceil \log_b C \rceil$ times, we reach a k-automatic sequence, say $(e_i)_{i \geq 0}$, whose terms are all $\leq b$, and $y \equiv \sum_{i \geq 0} e_i b^{-i-1} \pmod 1$.

The second step of the construction involves determining the carry bits that arise from the terms of e_i that equal b. Define the *carry sequence* $(c_i)_{i \geq 0}$ as follows:

$$c_i = \begin{cases} 1, & \text{if there exists } j > i \text{ with } x_{i+1} = x_{i+2} = \cdots = x_{j-1} = b - 1, \text{ and } x_j = b; \\ 0, & \text{otherwise.} \end{cases}$$

Then it is easy to see that if $f_i = ((e_i + c_i) \bmod b)_{i \geq 0}$, then $y \equiv \sum_{i \geq 0} f_i b^{-i-1} \pmod 1$, and $0 \leq f_i < b$. Thus it suffices to create a DFAO M that generates $(c_i)_{i \geq 0}$.

Our construction of $M = M_5$ goes in several stages. Let $M_0 = (Q, \Sigma, \delta, q_0, \Delta, \tau)$ be a DFAO generating $(e_i)_{i \geq 0}$; here $\Sigma = \{0, 1, \ldots, k-1\}$. First, we create a nondeterministic finite automaton (NFA) $M_1 = (Q', \Sigma \times \Sigma, \delta', q'_0, F)$ that, roughly speaking, has two non-negative integer inputs, i and j, and accepts if there exists n, $i < n < j$, such that $x_n \neq b - 1$. The inputs i and j are, of course, provided in base-k, with the shorter input padded by 0's in the front, if necessary, to make the lengths of the expansions the same. The NFA M_1 functions by nondeterministically guessing the base-k digits of n, and maintaining the relationship of the current guessed n with i and j.

The states of M_1 are triples of the form $[q, u, v]$, where $q \in Q$, and $u, v \in \{<, =\}$. The meaning of the state $[q, u, v]$ is that the guessed expansion of n seen so far would take us to state q in M_0, and furthermore the relationship of n to the currently seen inputs i and j is given by $i\ u\ n\ v\ j$ (e.g., $i < n = j$). The start state of M_1 is $q'_0 = [q_0, =, =]$. The transition function δ' is given by

$$\delta'([q, u, v], [c, d]) =$$

$$\begin{cases} [\delta(q, c), =, =], & \text{if } (u, v) = (=, =) \text{ and } c = d; \\ [\delta(q, c), =, <] \cup [\delta(q, d), <, =] \cup \bigcup_{c < z < d} [\delta(q, z), <, <], & \text{if } (u, v) = (=, =) \text{ and } c < d; \\ [\delta(q, d), <, =] \cup \bigcup_{0 \leq z < d} [\delta(q, z), <, <], & \text{if } (u, v) = (<, =); \\ [\delta(q, c), =, <] \cup \bigcup_{c < z < k} [\delta(q, z), <, <], & \text{if } (u, v) = (=, <); \\ \bigcup_{0 \leq z < k} [\delta(q, z), <, <], & \text{if } (u, v) = (<, <)\ . \end{cases}$$

Here c should be thought of as the next base-k digit of i; d should be thought of as the next digit of j, and z as the "guessed" next digit of n. The set of final states is given by

$$F = \{[q, <, <] \ : \ \tau(q) \neq b - 1\}.$$

We leave it to the reader to verify that M_1 really behaves as we have claimed.

Now, using the standard construction, we convert M_1 to a deterministic finite automaton (DFA) M_2 accepting the same set. Then, by interchanging accepting and non-accepting states of M_2, we get a DFA M_3 that accepts the base-k representations of pairs (i, j) such that for all n, with $i < n < j$, we have $e_n = b - 1$.

Next, we create a new NFA M_4 that, on input i, "guesses" the base-k digits of j and simulates M_3 on input (i, j). Our NFA M_4 also simulates M_0 on input j, and accepts iff M_3 accepts (i, j)

and M_0 outputs b on input j. Now M_4 can be easily converted to a DFAO M_5 that (essentially) generates the carry sequence $(c_i)_{i\geq 0}$. We say "essentially" because the base-k representation of j may have substantially more digits than that of i; hence only those base-k representations of i that have sufficiently many leading zeroes will result in the correct output. However, this problem may be easily dealt with using a trick of Eilenberg [4, Prop. 3.1, p. 106]. (Or see [14].) This completes the proof of the Lemma. ■

It remains to show that if $y, y' \in L(k,b)$, then $y + y' \in L(k,b)$. To see this, observe that $(y_i + y'_i)_{i\geq 0}$ is k-automatic, and then apply the Normalization Lemma to this sequence.

3 Dimension of $L(k,b)$ over $\mathbb{Q}$.

In the previous section, we saw that $L(k,b)$ is a vector space over $\mathbb{Q}$. Eric Bach asked (personal communication), what is the dimension of $L(k,b)$ and what is a basis? In this section, we answer the first question; the second is still open.

Theorem 2 *$L(k,b)$ is of infinite dimension over $\mathbb{Q}$.*

Proof. For simplicity, we prove the result for $k = 2$, but the proof can easily be modified to handle the general case.

Consider the formal power series

$$f(X) = \sum_{k\geq 0} X^{2^k} = X + X^2 + X^4 + X^8 + \cdots.$$

Then it is clear that, for all odd integers $r \geq 1$, the number $f(1/b^r)$ is a $(2,b)$-automatic real number, since a DFAO generating the base-b expansion of $f(1/b^r)$ need only output 1 if its input is of the form $(r)_2\, 0^*$, and output 0 otherwise.

We claim that the numbers $\{f(1/b^r) \; : \; r \text{ odd}, \; \geq 1\}$ are linearly independent over $\mathbb{Q}$. Assume not. Then there exists a finite linear combination

$$\sum_{0\leq i\leq s} a_i f(1/b^{2i+1}) = 0, \tag{1}$$

with $a_i \in \mathbb{Z}$ and not all $a_i = 0$. Let $M = \max_{0\leq i\leq s} |a_i|$.

Now separate the positive and negative coefficients in Eq. (1) to obtain a new equation

$$\sum_{0\leq i\leq s} d_i f(1/b^{2i+1}) = \sum_{0\leq i\leq s} e_i f(1/b^{2i+1}) \tag{2}$$

with $0 \leq d_i, e_i \leq M$ and $d_i e_i = 0$ for $0 \leq i \leq s$.

Now consider the base-b representation of both sides of Eq. (2). Let $(g)_b = g_0.g_1g_2g_3\cdots$ be the representation of the left-hand side, and $(h)_b = h_0.h_1h_2h_3\cdots$ be the representation of the right-hand side. These base-b representations are so sparse that for n large enough (it suffices to take $n > \lceil \log_2(2s+1)\rceil + \log_2(1+\log_b M)$), the digits immediately to the left of position $(2i+1)2^n$ in $(g)_b$ are $(d_i)_b$, while those in the same position in $(h)_b$ are $(e_i)_b$. It follows that $d_i = e_i$, and so $d_i = e_i = 0$ for $0 \leq i \leq s$. This gives us the desired contradiction. ■

4 An Infinite-Dimensional Automatic Ring $\mathbb{Q}[y]$

In the previous section, we proved that the vector space $L(k, b)$ is of infinite dimension over $\mathbb{Q}$ by exhibiting an infinite set of linearly independent automatic numbers. This raises the natural question, does there exist a *single* real number y whose positive powers are all automatic and linearly independent? In this section, we answer this question affirmatively. Again, for simplicity, we consider only the case $k = 2$, although our proof can be easily modified to handle the general case.

Theorem 3 *Let $y = f(1/b)$, where $f(X) = \sum_{k\geq 0} X^{2^k}$. Then every element in $\mathbb{Q}[y]$ is $(2, b)$-automatic, and furthermore $\mathbb{Q}[y]$ is of infinite dimension over $\mathbb{Q}$.*

Proof. Consider the number $y = f(1/b)$. Then $f(1/b)$ is transcendental, and hence the numbers $1, f(1/b), f(1/b)^2, f(1/b)^3, \ldots$ are linearly independent over $\mathbb{Q}$. This was first proved by Kempner [10]; for a more elementary proof, see [11]. Thus the result would follow if $f(1/b)^i$ were in $L(2, b)$ for $i \geq 2$.

To prove that $f(1/b)^i \in L(2, b)$ for $i \geq 2$, we use the theory of k-regular sequences. A sequence of integers $(c_n)_{n\geq 0}$ is said to be k-*regular* if its k-kernel

$$K_k(c) = \{(c_{k^r n+j})_{n\geq 0} \ : \ r \geq 0; \quad 0 \leq j < k^r\}$$

generates a finitely-generated module over $\mathbb{Z}$.

We now use the following theorems about k-regular sequences, as proved in [1]:

Theorem 4 *Every k-automatic sequence is k-regular.*

Theorem 5 *If $G(X) = \sum_{i\geq 0} g_i X^i$ and $H(X) = \sum_{i\geq 0} h_i X^i$ are both power series in $\mathbb{Z}[[X]]$, and their coefficient sequences $(g_i)_{i\geq 0}$ and $(h_i)_{i\geq 0}$ are both k-regular sequences, then so is the coefficient sequence of $G(X)H(X) = \sum_{i+j=n} g_i h_j X^n$.*

Theorem 6 *If a k-regular sequence is bounded, then it is k-automatic.*

Now define $\alpha(r, n) = [X^n] f(X)^r$, i.e., the coefficient of X^n in the formal power series $f(X)^r$. We now need the following Lemma, whose proof is postponed until the next section:

Lemma 7 *The quantity $\alpha(r, n)$ is bounded by a constant that depends on r, but not on n.*

It now follows from Theorems 4–6 and Lemma 7 that for any given r, the coefficients $(\alpha(r, n))_{n\geq 0}$ of $f(X)^r$ form an automatic sequence. Then, applying the Normalization Lemma, it follows that $f(1/b)^r$ is a $(2, b)$-automatic real number. Hence $\mathbb{Q}[y] \subseteq L(2, b)$. ■

5 $\alpha(r, n)$ is bounded.

The definition of $\alpha(r, n)$ given in the previous section implies the following interpretation: $\alpha(r, n)$ is the number of *compositions* of n as the sum of r integral powers of 2. (By "compositions" we mean that summands can be repeated and representations that differ only in the order of the summands are counted as distinct.)

To complete the proof of Theorem 3, we must prove Lemma 7: that $\alpha(r, n)$ is bounded by a constant that depends on r, but not on n.

Sections 5–7 of the paper are devoted to two proofs of this fact. Both lead to estimates on the size of $\alpha(r,n)$. The first provides a relationship with a previously-studied sequence counting the number of partitions of 1 as powers of 1/2, and leads to the estimate $\alpha(r,n) = O(r! \cdot 3.6^r)$. The second is inspired by notions of Kolmogorov complexity, and leads to a better bound of $\alpha(r,n) = O(r! \cdot 2^r)$.

In this section, we provide the first proof. We first show (Theorem 9) that any composition of n as the sum of powers of 2 can be "decomposed" into groups of terms, each of which sums to one of the powers of 2 appearing in the standard base-2 representation of n. It then suffices to bound $\alpha(r,n)$ where n is a power of 2. Next, we show (Lemma 11) that any composition of 2^j as the sum of r powers of 2 cannot include any terms smaller than 2^{j-r+1}. From this, we can conclude that $\alpha(r,n)$ is bounded for each r.

The claims in this section were obtained with the assistance of Anna Lubiw.

Lemma 8 *Let $x_0, x_1, x_2, \ldots, x_r$ be positive integers such $x_0 \geq x_1 \geq x_2 \geq \cdots \geq x_r$, $\sum_{1 \leq i \leq r} x_i \geq x_0$, and $x_{i+1} \mid x_i$ for $0 \leq i \leq r-1$. Then there exists an index b, $1 \leq b \leq r$, such that $\sum_{1 \leq i \leq b} x_i = x_0$.*

Proof. Omitted for space considerations. ∎

Theorem 9 *Let $n \geq 0$ be a positive integer, and express n as a sum of distinct powers of 2, i.e., $n = 2^{a_1} + 2^{a_2} + \cdots + 2^{a_t}$ where $a_1 < a_2 < \cdots < a_t$. Then if there exists a multiset S of non-negative integers such that $n = \sum_{s \in S} 2^s$, then S can be partitioned into t disjoint submultisets $S_1, S_2, \ldots, S_t$ such that $2^{a_i} = \sum_{s \in S_i} 2^s$ for $1 \leq i \leq t$.*

Proof. By induction on t. The details will appear in the final paper. ∎

Lemma 10 *Define $s_2(n)$ to be the sum of the bits in the base-2 expansion of n. Then if there exist non-negative integers $c_1, c_2, \ldots, c_j$ (not necessarily distinct) such that $n = 2^{c_1} + 2^{c_2} + \cdots + 2^{c_j}$, we must have $j \geq s_2(n)$.*

Proof. Clear. ∎

Lemma 11 *Suppose $c_1, c_2, \ldots, c_j, m$ are non-negative integers such that $c_1 \leq c_2 \leq \cdots \leq c_j$ and $2^m = 2^{c_1} + \cdots + 2^{c_j}$. Then $c_1 \geq m - j + 1$.*

Proof. We have $2^m - 2^{c_1} = 2^{c_2} + \cdots + 2^{c_j}$, and $s_2(2^m - 2^{c_1}) = m - c_1$. By Lemma 10, we have $j - 1 \geq m - c_1$, and the result follows. ∎

Lemma 12 *Let $U(r,m)$ denote the number of compositions of 2^m as a sum of r powers of 2. Then $U(r,m) \leq r^r$.*

Proof. By Lemma 11, we know that any such composition of 2^m must use only the powers $2^m, 2^{m-1}, \ldots, 2^{m-r+1}$. Thus, at most r different powers of 2 can be used, and each power might potentially appear in any one of r different places. This gives the bound. ∎

Note that the bound $U(r,m) \leq r^r$ may be easily improved to $(r-1)^r$ for $r \geq 2$, by observing that 2^m cannot be used in any composition of 2^m as a sum of 2 or more powers of 2.

We can now complete the first proof of Lemma 7, and prove that $\alpha(r,n)$ is bounded. If $s_2(n) > r$, then by Lemma 10, it follows that $\alpha(r,n) = 0$. If $s_2(n) \leq r$, then by Theorem 9 we can express n in base 2, say $n = 2^{a_1} + \cdots + 2^{a_t}$, and consider separately the composition of each 2^{a_i}. By Lemma 12, there are at most r^r compositions for each 2^{a_i}, and since $t \leq r$, there are only a finite number of compositions of n as the sum of r powers of 2. ∎

6 Representations as sums of powers of two.

In the previous section, we introduced $U(r,m)$, the number of compositions of 2^m as a sum of r powers of 2.

Numerical evidence suggests that $U(r,m)$ is eventually constant, as m gets large. This is clearly true, since for $m \geq r-1$, any composition of 2^m corresponds in a 1-1 fashion with a composition for $2^{m'}$, $(m' > m)$, by multiplication by the appropriate power of 2. This suggests defining $U_r = U(r, r-1)$.

We may improve the result of Lemma 12 still further by studying the unordered analogue of $U(r,m)$. Let $V(r,m)$ denote the number of *partitions* of 2^m as a sum of r powers of 2. (By "partition" we mean that summands may be repeated, but representations that differ only in order of the summands are regarded as identical.) Then clearly $U(r,m) \leq r!V(r,m)$.

We now relate $V(r,m)$ to a quantity that has been previously studied by many authors: namely, H_r, the number of partitions of 1 as the sum of r powers of 1/2. Such a partition of 1 gives rise (by multiplication by an appropriate power of 2) to a partition of 2^m as a sum of r powers of 2. The converse also holds. It therefore follows that $V(r,m) \leq H_r$, and indeed $V(r,m) = H_r$ for $m \geq r-1$. (This idea also suggests another way of expressing U_r, as the ordered analogue of H_r: the number of compositions of 1 as a sum of r powers of 1/2.)

Many other authors have studied the sequence

$$(H_r)_{r\geq 1} = (1,1,1,2,3,5,9,16,28,50,89,\ldots),$$

which arises in computing prefix codes for trees [5]; enumeration of elements in groupoids [13]; enumeration of codes [8]; and algebraic topology [2, 6, 16]. It is Sloane's sequence #261 [15].

It is known [6] that $H_r \sim K \cdot \nu^{r-1}$, where $K \doteq .25451$ and $\nu \doteq 1.79415$. Hence we have the following improved bound for U_r and $U(r,m)$:

Theorem 13 *There exists a constant K' such that $U(r,m) \leq U_r \leq K' \cdot r! \cdot 1.8^r$.*

It would be of interest to determine the asymptotic behaviour of U_r. Numerical evidence suggests that perhaps $U_r \sim A \cdot r! \cdot B^r$, where $A \doteq .2487$ and $B \doteq 1.1926$. (N. J. A. Sloane was kind enough to send us a copy of a letter dated July 22, 1975 from D. E. Knuth to R. E. Tarjan. In this letter Knuth studies U_r and suggests that "something like" $U_r \sim c_1 r^{r-c_2}$ should be true for constants c_1, c_2.)

Now define $W_r = \max_{n\geq 0} \alpha(r,n)$. We have

Theorem 14 *There exists a constant K_2 such that $W_r \leq K_2 \cdot r! \cdot 3.6^r$.*

Proof. It follows from the proof of Lemma 7 that $\alpha(r,n)$ achieves its maximum when the base-2 expansion of n is of the form

$$(n)_2 = (10^{r-1})^i = \underbrace{(1\overbrace{00\cdots0}^{r-1})\cdots(1\overbrace{00\cdots0}^{r-1})}_{i}$$

for some i, $1 \leq i \leq r$. (It may also achieve this maximum at other strings.) For by Theorem 9, we may consider separately the representations for each power of 2 in the binary expansion of n, and by Lemma 11, the representation for 2^n uses only the terms $2^n, 2^{n-1}, \ldots, 2^{n-r+1}$. To maximize $\alpha(r,n)$, we can assume that the ranges of the representations for the various powers of 2 that appear in the binary representation of n do not overlap; for duplicate occurrences of the same power of 2 would lead to fewer compositions. This gives us a way to estimate W_r, using the previously cited bound for H_r.

There are $\binom{r-1}{i-1}$ compositions of r as the sum of i positive integers, $r = b_1 + b_2 + \cdots + b_i$. For each such composition, we can partition $2^{j(r-1)}$ $(1 \le j \le i)$ into b_j powers of 2 in H_{b_j} different ways. Finally, once an unordered representation for n is chosen, it may be re-ordered in at most $r!$ ways. This gives the bound

$$\begin{aligned} W_r &\le r! \max_{1\le i\le r} \binom{r-1}{i-1} \max_{\substack{b_1+\cdots+b_i=r\\ b_1,\ldots,b_i\ge 1}} \prod_{1\le j\le i} H_{b_j} \\ &\le r! \max_{1\le i\le r} \binom{r-1}{i-1} \cdot K_1 \cdot 1.8^r \\ &\le r! \cdot \frac{2^{r-1}}{\sqrt{\pi(r-1)/2}} \cdot K_1 \cdot 1.8^r \\ &\le K_2 \cdot r! \cdot 3.6^r, \end{aligned}$$

where K_1, K_2 are constants. ■

It would be of interest to determine the true asymptotic behaviour of W_r. Numerical evidence suggests that perhaps $W_r \sim C \cdot r! \cdot D^r$, where $C \doteq .131$ and $D \doteq 1.686$.

7 An improved bound on $\alpha(r, n)$.

In this section, we give another proof of the fact that $\alpha(r, n)$ is bounded, for each fixed r. This proof provides a better estimate for $\alpha(r, n)$.

The idea of the proof is to encode each sequence of r powers of 2 adding up to n as a pair of sequences characterizing the additions in binary notation. Suppose $n = 2^{j_1} + 2^{j_2} + \cdots + 2^{j_r}$. Define $n_i = 2^{j_1} + \cdots + 2^{j_i}$, the ith partial sum, and consider the addition $n_i = n_{i-1} + 2^{j_i}$ in base-2:

$$\begin{array}{cccccc} x_i & 0 & 1^{d_i} & y_i & = & n_{i-1} \\ + & & 1 & 0^{j_i} & = & 2^{j_i} \\ \hline x_i & 1 & 0^{d_i} & y_i & = & n_i \end{array} \tag{3}$$

Define a_i to be the number of 1's in the string x_i, and $b_i = s_2(n_i)$, the number of 1's in the binary expansion of n_i. Then

$$0 \le a_i < b_i, \tag{4}$$

$$1 \le b_i \le b_{i-1} + 1. \tag{5}$$

Given n_i, the addition (3) is completely determined by a_i (which determines the 1 where carry propagation ends) and b_{i-1} (which gives the "carry distance" d_i as $b_{i-1} + 1 - b_i$). It follows that the sequence $(2^{j_1}, 2^{j_2}, \ldots, 2^{j_r})$ is completely characterized by the pair of sequences

$$\begin{array}{lllll} a_1 = 0, & a_2, & \ldots, & a_{r-1}, & a_r; \\ b_1 = 1, & b_2, & \ldots, & b_{r-1}, & b_r. \end{array}$$

We can relate these (a, b)-sequences to the *Catalan Tree* T, defined as follows: the root of T is labeled 1, and each vertex labeled i has $i + 1$ children labeled $1, 2, \ldots, i + 1$. (This tree, up to a relabeling, was studied previously by J. West [18, 17, 19], who observed that the number of vertices

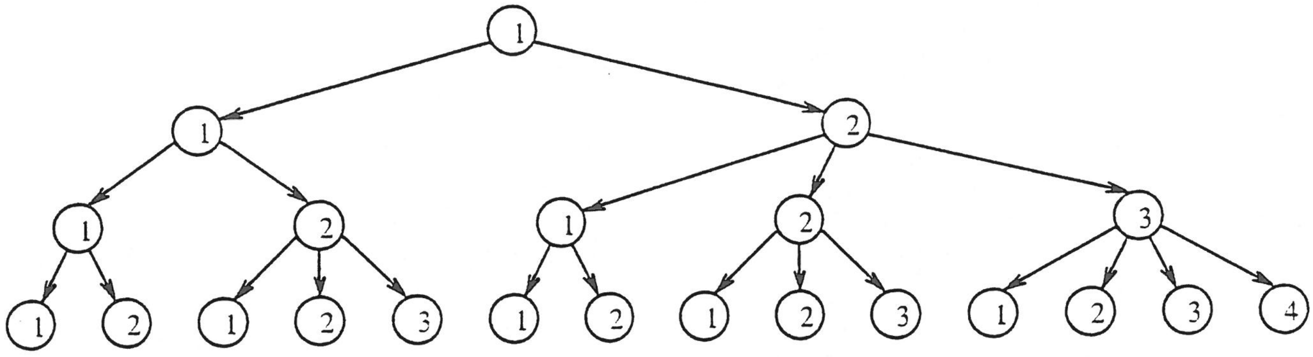

Figure 1: Levels 0–3 of the Catalan Tree T

at depth d is the $d+1$st Catalan number, $\binom{2d+2}{d+1}/(d+2)$.) Let the *weight* of a vertex v be the product of all labels on the path from the root to v, and let $w(d,i)$ be the sum of the weight of all vertices labeled i at depth d.

Levels 0–3 of T are given in Fig. 1. Note that each sequence $b_1,\ldots,b_r$ satisfying (5) corresponds to a path from the root to a vertex v at depth $r-1$, and the weight of the vertex v equals the number of possible sequences $a_1,\ldots,a_r$ satisfying (4). This gives us the upper bound $\alpha(r,n) \leq \max_{1\leq j\leq r} w(r-1,j) \leq w(r,1)$. The bound is not sharp; for example, only 2 of the 4 a-sequences are feasible for the b-sequence $1,2,2,1$. The reason is that the two 1's in n_3 must be adjacent to allow a carry distance of 2 from n_3 to n_4, hence the carry from n_2 to n_3, having distance 1, can only go into one of these 1's.

We can compute $w(d,i)$ by defining $w(d,0)=0$ for all d, and using the following recurrence relation:

$$w(d,i) = i \sum_{i-1\leq j\leq d} w(d-1,j). \tag{6}$$

We now prove the following formula for $w(d,i)$:

Theorem 15 *We have*

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

Proof. First, we prove the following lemma:

Lemma 16 *For all integers $n, d \geq 0$ we have*

$$\sum_{0\leq j\leq n} j\cdot 2^j \cdot \binom{2d-j-1}{d-1} = d\binom{2d}{d} - d\cdot 2^{n+1}\cdot \binom{2d-n-1}{d}.$$

Proof. An easy induction on n. The details will be given in the final paper. ■

Now we can prove Theorem 15, by induction on d. The base case, $d=0$, is left to the reader. Assume the result is true for $d-1$. Then we have

$$w(d,i) \;=\; i \sum_{i-1\leq j\leq d} w(d-1,j)$$

$$
\begin{aligned}
&= i \sum_{i-1 \le j \le d} \frac{j(2d-j-1)!}{2^{d-j}(d-j)!} \qquad \text{(by induction)} \\
&= \frac{i(d-1)!}{2^d} \sum_{i-1 \le j \le d} j \cdot 2^j \cdot \binom{2d-j-1}{d-1} \\
&= \frac{i(d-1)!}{2^d} \cdot d \cdot 2^{i-1} \cdot \binom{2d-i+1}{d} \qquad \text{(by Lemma 16)} \\
&= \frac{i(2d-i+1)!}{2^{d-i+1}(d-i+1)!}.
\end{aligned}
$$

That completes the proof. ■

By setting $i = 1$, it now follows that $w(d,1) = (2d)!/(2^d \cdot d!)$. (There is also a beautiful combinatorial proof of this fact, which is based on a 1-1 correspondence between (a,b)-sequences and perfect matchings on the complete graph on $2d$ labeled vertices.)

We have therefore proved the following theorem:

Theorem 17 $W_r = \max_{n \ge 0} \alpha(r,n)$ *exists and we have*

$$W_r \le w(r,1) = \frac{(2r)!}{2^r r!} \sim 2^r \cdot r! \cdot (\pi r)^{-1/2}.$$

8 $L(k,b)$ is not closed under product.

At the "Thémate" conference in Luminy in May 1993, Lehr raised the question of whether or not the automatic real numbers form a ring, i.e., are they closed under product? In this section we resolve this open problem by exhibiting a counterexample. It follows that $L(k,b)$ is not closed under squaring or taking the reciprocal.

Theorem 18 $L(k,b)$ *is not closed under product.*

Proof. For simplicity we prove the result only for $k = 2$, although the methods can easily be extended to cover the general case. As before, we define

$$f(X) = \sum_{r \ge 0} X^{2^r} = X + X^2 + X^4 + \cdots.$$

We also define

$$
\begin{aligned}
g(X) &= \sum_{m \ge 1,\ n \ge 0} X^{(2^m-1)2^n} \\
&= X + X^2 + X^3 + X^4 + X^6 + X^7 + X^8 + X^{12} + X^{14} + X^{15} + X^{16} + X^{24} + X^{28} + X^{30} + \cdots
\end{aligned}
$$

As before, if $y = f(1/b)$, then $y \in L(2,b)$. Similarly, if $z = g(1/b)$, then $z \in L(2,b)$, since the base-b representation of z has 1's in those positions whose base-2 representations are given by the regular set 1^+0^*. We will show that $yz \notin L(2,b)$.

First, note that

$$
\begin{aligned}
f(X)g(X) &= \sum_{m \ge 1,\ n \ge 0,\ r \ge 0} X^{2^r} \cdot X^{(2^m-1)2^n} = \sum_{m \ge 1,\ n \ge 0,\ r \ge 0} X^{2^r+(2^m-1)2^n} \\
&= \sum_{\substack{r<n \\ m \ge 1,\ n \ge 0,\ r \ge 0}} X^{2^r+(2^m-1)2^n} + \sum_{\substack{r=n \\ m \ge 1,\ n \ge 0,\ r \ge 0}} X^{2^r+(2^m-1)2^n} + \sum_{\substack{r>n \\ m \ge 1,\ n \ge 0,\ r \ge 0}} X^{2^r+(2^m-1)2^n} \\
&= S(X) + T(X) + U(X).
\end{aligned}
$$

Second, note that

$$\begin{aligned} S(X) &= \sum_{\substack{r<n \\ m\geq 1,\ n\geq 0,\ r\geq 0}} X^{2^r+(2^m-1)2^n} = \sum_{\substack{r<n \\ m\geq 1,\ n\geq 0,\ r\geq 0}} X^{2^r(1+2^{n-r}(2^m-1))} \\ &= \sum_{m\geq 1,\ p\geq 1,\ r\geq 0} X^{2^r(1+2^p(2^m-1))} = \sum_{m\geq 1,\ p\geq 1,\ r\geq 0} X^{2^r(2^{p+m}-2^p+1)}. \end{aligned}$$

Now $(2^r(2^{p+m}-2^p+1))_2 = 1^m\ 0^{p-1}\ 1\ 0^r$, so it follows that $S(X) = \sum_{i\geq 0} s_i X^i$, where

$$s_i = \begin{cases} 1, & \text{if } (i)_2 \in 1^+\ 0^*\ 1\ 0^*; \\ 0, & \text{otherwise.} \end{cases}$$

Hence $(s_i)_{i\geq 0}$ is a 2-automatic sequence, and therefore $S(1/b) \in L(2,b)$.

By the same technique, it can be shown that $U(x) = \sum_{i\geq 0} u_i X^i$, where

$$u_i = \begin{cases} 2, & \text{if } (i)_2 \in 1\ 0^+\ 1^+\ 0^*; \\ 1, & \text{if } (i)_2 \in 1\ 1^+\ 0^*; \\ 0, & \text{otherwise.} \end{cases}$$

It follows that $(u_i)_{i\geq 0}$ is a 2-automatic sequence, and, by the Normalization Lemma, we have $U(1/b) \in L(2,b)$.

Finally, note that

$$T(X) = \sum_{\substack{r=n \\ m\geq 1,\ n\geq 0,\ r\geq 0}} X^{2^r+(2^m-1)2^n} = \sum_{r\geq 0,\ m\geq 1} X^{2^r(1+2^m-1)} = \sum_{r\geq 0,\ m\geq 1} X^{2^{m+r}} = \sum_{n\geq 1} nX^{2^n}.$$

Now consider the base-b expansion of $T(1/b)$, say $(T(1/b))_b = 0.c_0c_1c_2\cdots$. Evidently the base-$b$ digits immediately to the left of position 2^n are just $(n)_b$. It follows that every element of $\{0,1,\ldots,b-1\}^*$ eventually appears as a factor of the infinite sequence

$$c = c_0c_1c_2\ldots.$$

If we define the subword complexity $p_d(n)$ of an infinite sequence $d = (d_i)_{i\geq 0}$ to be the number of distinct factors of length n which appear in d, then we have shown that $p_c(n) = b^n$. But, by a result of Cobham [3], if c were k-automatic for any k, we would have $p_c(n) = O(n)$. This gives a contradiction, and so $T(1/b) \notin L(2,b)$.

It follows that $yz \notin L(2,b)$, since $yz = S(1/b) + T(1/b) + U(1/b)$. ■

Theorem 19 *The set $L(2,b)$ is not closed under the map $x \to x^2$.*

Proof. Suppose it were. Then, since $yz = \frac{1}{4}((y+z)^2 - (y-z)^2)$, we would have that $L(2,b)$ is closed under product, a contradiction. ■

Actually, it can be shown (using the same techniques as in the proof of Theorem 18), that $g(1/b)^2 \notin L(2,b)$.

Theorem 20 *The set $L(2,b)$ is not closed under the map $x \to 1/x$.*

Proof. Suppose it were. Then, since $y^2 = y + \frac{1}{\frac{1}{y-1}-\frac{1}{y}}$, we would have that $L(2,b)$ is closed under squaring, a contradiction. ■

9 Acknowledgments.

We thank Eric Bach for having suggested the problem in Section 3, Christiane Frougny for informing us about Eilenberg's trick, and Neil Sloane for sending us a copy of the Knuth-Tarjan correspondence. Julian West kindly sent us a copy of some of his unpublished work. We also thank Anna Lubiw, Jean-Paul Allouche, Bruce Reznick, Bruce Richmond and Alfredo Viola for useful discussions. Thanks also to the anonymous, strategic referee.

References

[1] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.* **98** (1992), 163–187.

[2] G. Carlsson. G. B. Segal's Burnside ring conjecture for $(\mathbb{Z}/2)^k$. *Topology* **22** (1983), 83–103.

[3] A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.

[4] S. Eilenberg. *Automata, Languages, and Machines*, Vol. A. Academic Press, San Diego, 1974.

[5] S. Even and A. Lempel. Generation and enumeration of all solutions of the characteristic sum condition. *Inform. Control* **21** (1972), 476–482.

[6] P. Flajolet and H. Prodinger. Level number sequences for trees. *Discrete Math.* **65** (1987), 149–156.

[7] C. Frougny. Representations of numbers and finite automata. *Math. Systems Theory* **25** (1992), 37–60.

[8] E. N. Gilbert. Codes based on inaccurate source probabilities. *IEEE Trans. Inform. Theory* **IT-17** (1971), 304–314.

[9] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.

[10] A. J. Kempner. On transcendental numbers. *Trans. Amer. Math. Soc.* **17** (1916), 476–482.

[11] M. J. Knight. An "ocean of zeros" proof that a certain non-Liouville number is transcendental. *Amer. Math. Monthly* **98** (1991), 947–949.

[12] S. Lehr. Sums and rational multiples of q–automatic sequences are q–automatic. *Theoret. Comput. Sci.* **108** (1993), 385–391.

[13] H. Minc. A problem in partitions: enumeration of elements of a given degree in the free commutative entropic cyclic groupoids. *Proc. Edinburgh Math. Soc.* **11** (1959), 223–224.

[14] J. Shallit and D. Wilson. The $3x + 1$ problem and finite automata. *Bull. European Assoc. Theor. Comput. Sci.* (46) (February 1992), 182–185.

[15] N. J. A. Sloane. *A Handbook of Integer Sequences*. Academic Press, New York, 1973.

[16] M. C. Tangora. Level number sequences of trees and the lambda algebra. *Europ. J. Combinatorics* **12** (1991), 433–443.

[17] J. West. Generating trees and the Catalan and Schröder numbers. *Discrete Math.*, to appear.

[18] J. West. *Permutations with forbidden subsequences; and, stack-sortable permutations.* PhD thesis, MIT, 1990. Department of Mathematics.

[19] J. West. Generating trees and forbidden subsequences. In *Proc. 6th Conference on Formal Power Series and Algebraic Combinatorics*, pp. 441–450. 1994.