PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY, SERIES B
Volume 5, Pages 50–63 (August 2, 2018)
https://doi.org/10.1090/bproc/37

WHEN IS AN AUTOMATIC SET AN ADDITIVE BASIS?

JASON BELL, KATHRYN HARE, AND JEFFREY SHALLIT

(Communicated by Matthew A. Papanikolas)

Abstract. We characterize those k-automatic sets S of natural numbers that form an additive basis for the natural numbers, and we show that this characterization is effective. In addition, we give an algorithm to determine the smallest j such that S forms an additive basis of order j, if it exists.

1. Introduction

One of the principal problems of additive number theory is to determine, given a set $S \subseteq \mathbb{N}$, whether there exists a constant j such that every natural number (respectively, every sufficiently large natural number) can be written as a sum of at most j members of S (see, e.g., [23]). If such a j exists, we say that S is an *additive basis* (resp., an *asymptotic additive basis*) *of order* j for $\mathbb{N}$.

Variants of this problem date back to antiquity, with Diophantus asking whether every natural number could be expressed as a sum of four squares. More generally, Waring's problem asks whether the set of k-th powers forms an additive basis for the natural numbers, which was ultimately answered in the affirmative by Hilbert [23, Chapter 3]. The problem of finding bounds on the number of k-th powers required to express all natural numbers and all sufficiently large natural numbers, as well as whether restricted subsets of k-th powers form additive bases, continues to be an active area of research [30–32].

Independent of Hilbert's work on Waring's problem, the famed Goldbach conjecture asks whether every even positive integer can be expressed as the sum of at most two prime numbers. If true, this would then imply that every sufficiently large natural number is the sum of at most three prime numbers. Vinogradov [23, Chapter 8] has shown that every sufficiently large natural number can be expressed as the sum of at most four prime numbers, and so the set of prime numbers is an asymptotic additive basis for the natural numbers.

From these classical beginnings, a general theory of additive bases has since emerged, and the problem of whether given sets of natural numbers form additive bases (or asymptotic additive bases) has been considered for many classes of sets.

If one adopts a computational point of view, subsets of natural numbers can be divided into two classes: computable sets (i.e., sets that can be produced using a Turing machine) and those sets that lie outside the realm of classical computation.

Received by the editors October 23, 2017, and, in revised form, March 27, 2018.
2010 *Mathematics Subject Classification.* Primary 11B13; Secondary 11B85, 68Q45, 28A80.
Key words and phrases. Additive basis, automatic set, finite-state automaton, Cantor sets.
Research of the first author was supported by NSERC Grant 2016-03632.
Research of the second author was supported by NSERC Grant 2016-03719.
Rsearch of the third author was supported by NSERC Grant 105829/2013.

Historically, the explicitly given sets for which the problem of being an additive basis has been considered are computable, and a natural problem is to classify the computable subsets of the natural numbers that form additive bases. However, a classical theorem of Kreisel, Lacombe, and Shoenfield [16] implies that the question of whether a given computable subset of $\mathbb{N}$ forms an additive basis is, in general, recursively unsolvable. Even for relatively simple sets, the problem seems intractable, as it applies to many sets of natural numbers, such as the set of twin primes, for which it is still open as to whether it is infinite, let alone whether it is an additive basis, which heuristics indicate should be the case [33]. Thus it is of interest to identify some classes of sets for which the problem is decidable.

One mechanism for producing computable sets is to fix a natural number $k \geq 2$ and consider natural numbers in terms of their base-k expansions. A set of natural numbers can then be regarded as a sublanguage of the collection of words over the alphabet $\{0, 1, \ldots, k-1\}$. In this setting, there is a coarse hierarchy, formulated by Chomsky, that roughly divides complexity into four nested classes: recursively enumerable languages (those that are produced using Turing machines); context-sensitive languages (those produced using linear-bounded non-deterministic Turing machines); context-free languages (those produced using pushdown automata); and regular languages (those produced using finite-state automata). The simplest of these four classes is the collection of regular languages. When one uses a regular sublanguage of the collection of words over $\{0, 1, \ldots, k-1\}$, the corresponding collection of natural numbers one obtains is called a k-*automatic set* (see, for example, [2]).

In this paper we completely characterize those k-automatic sets of natural numbers that form an additive basis or an asymptotic additive basis. In the case of a k-automatic set S of natural numbers, there is a well-understood dichotomy: either $\pi_S(x) := \#\{n \leq x \colon n \in S\}$ is $O((\log x)^d)$ for some natural number d or there is a real number $\alpha > 0$ such that $\pi_S(x) = \Omega(x^\alpha)$ (see Section 2 and specifically Corollary 2.7 for details). In the case where $\pi_S(x)$ is asymptotically bounded by a power of $\log x$, we say that S is *sparse*. Our first main result is the following theorem (see Theorem 4.1 and the remarks that follow).

Theorem 1.1. *Let $k \geq 2$ be a natural number and let S be a k-automatic subset of $\mathbb{N}$. Then S forms an asymptotic additive basis for $\mathbb{N}$ if and only if the following conditions both hold:*

(1) *S is not sparse;*
(2) $\gcd(S) = 1$.

Moreover, if S is a non-sparse set and $\gcd(S) = 1$, then there exist effectively computable constants M and N such that every natural number greater than or equal to M can be expressed as the sum of at most N elements of S.

We note that a necessary condition for a set S to be an additive basis is that 1 be in S. If S is not sparse and $\gcd(S) = 1$ and $1 \in S$, then S is an additive basis, and these conditions are necessary. We give explicit upper bounds on M and N in terms of the number of states in the minimal automaton that accepts the set S, and we show that these bounds are in some sense the correct form for the type of bounds one expects to hold in general. An interesting feature of our proof is that it uses results dealing with sums of Cantor sets obtained by the second-named author in work with Cabrelli and Molter [7].

Our second main result is the following.

Theorem 1.2. *Let $k \geq 2$ be a natural number and let S be a k-automatic subset of $\mathbb{N}$. There is an algorithm that determines whether the conditions of Theorem* 1.1 *hold and, if so, also determines the smallest possible N in that theorem and the corresponding smallest possible M.*

The outline of this paper is as follows. In Section 2 we recall some of the basic concepts from the theory of regular languages and automatic sets, including the notion of a sparse automatic set, which play a key role in the statement of Theorem 1.1. In Section 3 we give some of the necessary background on Cantor sets and prove a key lemma involving these sets. In Section 4 we prove a strengthening of Theorem 1.1 (see Theorem 4.1) that gives explicit bounds on M and N appearing in the statement of the theorem. In Section 5, we give an algorithm that allows one to find optimal bounds for given automatic sets, and in Section 6, we give some examples to illustrate the usage of our algorithm.

For other recent results connecting additive number theory and formal language theory, see [15, 19, 24, 25].

2. Basics

We are concerned with words and numbers. A *word* is a finite string of symbols over a finite alphabet Σ. If x is a word, then $|x|$ denotes its length (the number of symbols in it). The *empty word* is the unique word of length 0, and it is denoted by ϵ.

The *canonical base-k expansion* of a natural number n is the unique word over the alphabet $\Sigma_k = \{0, 1, \ldots, k-1\}$ representing n in base k, without leading zeros, starting with the most significant digit. It is denoted $(n)_k$. Thus, for example, $(43)_2 = 101011$. If w is a word, possibly with leading zeros, then $[w]_k$ denotes the integer that w represents in base k.

A *language* is a set of words. Three important languages are

(i) Σ^*, the set of all finite words over the alphabet Σ;
(ii) Σ^n, the set of words of length n; and
(iii) $\Sigma^{\leq n}$, the set of words of length $\leq n$.

Given a set $S \subseteq \mathbb{N}$, we write $(S)_k$ for the language of canonical base-k expansions of elements of S.

There is an ambiguity that arises from the direction in which base-k expansions are read by an automaton. In this article we always assume that these expansions are read starting with the least significant digit.

We recall the standard asymptotic notation for functions from $\mathbb{N}$ to $\mathbb{N}$:

- $f = O(g)$ means that there exist constants $c > 0$, $n_0 \geq 0$ such that $f(n) \leq cg(n)$ for $n \geq n_0$;
- $f = \Omega(g)$ means that there exist constants $c > 0$, $n_0 \geq 0$ such that $f(n) \geq cg(n)$ for $n \geq n_0$;
- $f = \Theta(g)$ means that $f = O(g)$ and $f = \Omega(g)$.

Given a language L defined over an alphabet Σ, its *growth function* $g_L(n)$ is defined to be $|L \cap \Sigma^n|$, the number of words in L of length n. If there exists a real number $\alpha > 1$ such that $g_L(n) > \alpha^n$ for infinitely many n, then we say that L has *exponential growth*. If there exists a constant $c \geq 0$ such that $g_L(n) = O(n^c)$, then we say that L has *polynomial growth*.

A *deterministic finite automaton* or DFA is a quintuple $M = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite non-empty set of states, Σ is the input alphabet, q_0 is the initial state, $F \subseteq Q$ is a set of final states, and $\delta : Q \times \Sigma \to Q$ is the transition function. The function δ can be extended to $Q \times \Sigma^* \to Q$ in the obvious way. The language accepted by M is defined to be $\{x \in \Sigma^* \ : \ \delta(q_0, x) \in F\}$. A language is said to be *regular* if there is a DFA accepting it [13].

A *non-deterministic finite automaton* or NFA is like a DFA, except that the transition function δ maps $Q \times \Sigma$ to 2^Q. A word x is accepted if some path labeled x causes the NFA to move from the initial state to a final state.

We now state three well-known results about the growth functions of regular languages. These lemmas follow by combining the results in, e.g., [9, 10, 14, 28, 29].

Lemma 2.1. *Let L be a regular language. Then L has either polynomial or exponential growth.*

Define $h_L(n) = |L \ \cap \ \Sigma^{\leq n}|$, the number of words of length $\leq n$.

Lemma 2.2. *Let L be a regular language. The following are equivalent:*

(a) *L is of polynomial growth;*
(b) *there exists an integer $d \geq 0$ such that $h_L(n) = \Theta(n^d)$;*
(c) *L is the finite union of languages of the form $z_0 x_1^* z_1 x_2^* \cdots z_{i-1} x_i^* z_i$ for words $z_0, z_1, \ldots, z_i$, $x_1, x_2, \ldots, x_i$;*
(d) *there exist a constant j and words $y_1, y_2, \ldots, y_j$ such that $L \subseteq y_1^* y_2^* \cdots y_j^*$.*

Lemma 2.3. *Let L be a regular language, accepted by a DFA or NFA $M = (Q, \Sigma, \delta, q_0, F)$. The following are equivalent:*

(a) *L is of exponential growth;*
(b) *there exists a real number $\rho > 1$ such that $h_L(n) = \Omega(\rho^n)$;*
(c) *there exists a state q of M and words w_0, x_0, x_1, z_0 such that $x_0 x_1 \neq x_1 x_0$ and $\delta(q_0, w_0) = \delta(q, x_0) = \delta(q, x_1) = q$, and $\delta(q, z_0) \in F$;*
(d) *there exist words w, x, y, z with $xy \neq yx$ such that $w\{x, y\}^* z \subseteq L$;*
(e) *there exist words s, t, u, v with $|t| = |u|$ and $t \neq u$ such that $s\{t, u\}^* v \subseteq L$.*

We will also need the following result, which appears to be new.

Lemma 2.4. *In Lemma* 2.3(e)*, the words s, t, u, v can be taken to obey the following inequalities: $|s|, |v| < n$ and $|t|, |u| < 3n$, where n is the number of states in the smallest DFA or NFA M accepting L.*

Proof. Consider those quadruples of words (w_0, x_0, x_1, z_0) satisfying the conditions of Lemma 2.3(c), namely, that there is a state q of M such that $\delta(q_0, w_0) = \delta(q, x_0) = \delta(q, x_1) = q$, and $\delta(q, z_0) \in F$, and $x_0 x_1 \neq x_1 x_0$. We can choose w_0 and z_0 minimal so that no state is encountered more than once via the paths P_{w_0} and P_{z_0} through M labeled w_0 and z_0, respectively. Thus without loss of generality we can assume $|w_0|, |z_0| < n$.

Next, among all such x_0, x_1, assume x_0 is a shortest non-empty word and x_1 is a shortest non-empty word paired with x_0. Consider the set of states encountered when going from q to q via the path P_{x_0} labeled x_0. If some state (other than q) is encountered twice or more, this means there is a loop we can cut out and find a shorter non-empty word x_0' with $\delta(q, x_0') = q$. By minimality of the length of x_0, we must have that x_0' commutes with all words w such that $\delta(q, w) = w$. In particular, x_0' commutes with x_0 and x_1. Since the collection of words that commute with a non-trivial word consists of powers of a common word [18, Proposition 1.3.2], we see that if this were the case, then x_0 and x_1 would commute, a contradiction. Thus $|x_0| \leq n$. By construction $|x_1| \geq |x_0|$. If x_0 is a proper prefix of x_1, then we have $x_1 = x_0 x_1'$ for some non-empty word x_1' with $\delta(q, x_1') = q$, and since $x_0 x_1 \neq x_1 x_0$, we have $x_0 x_0 x_1' \neq x_0 x_1' x_0$. Cancelling x_0 on the left gives $x_0 x_1' \neq x_1' x_0$. But this contradicts minimality of the length of x_1.

Thus x_1 has some prefix p with $|p| \leq |x_0|$ such that $x_1 = pp'$ and p is not a prefix of x_0. Let $q' = \delta(q, p)$. If $q' = q$, then we have $\delta(q, p) = q$, and $x_0 p \neq p x_0$, since p is not a prefix of x_0. Thus in this case, by minimality of x_1, we have $x_1 = p$ and so $|x_1| \leq n$. Thus we may assume that $q' \neq q$. Then $\delta(q', p') = q$. Let u be the label of a shortest path from q' to q. Then $|u| < n$ since by removing loops, we may assume the path P_u visits no state more than once and it does not revisit q'. Observe that $|pu| < 2n$ and $\delta(q, pu) = q$. Moreover, $x_0 pu \neq pu x_0$ since p is not a prefix of x_0. Thus, by the minimality of x_1, we have $|x_1| \leq |up| < 2n$.

Thus we can assume that $|x_0| \leq n$ and $|x_1| < 2n$. Setting $s = w_0$, $t = x_0 x_1$, $u = x_1 x_0$, and $v = z_0$ gives the desired inequalities. □

Remark 2.5. The bound $3n - 1$ in Lemma 2.4 is optimal. For example, consider an NFA $M = (\{q_1, \ldots, q_n\}, \{a, b\}, \delta, q_1, \{q_1\})$ with n states $q_1, q_2, \ldots, q_n$ connected in a directed cycle with transitions labeled by a. Add a directed edge labeled b from q_n back to q_2. Then the smallest words obeying the conditions are $x = a^n$ of length n and $y = a^{n-1} b a^{n-1}$ of length $2n - 1$. Then $t = xy$ and $u = yx$ and $|t| = |u| = 3n - 1$.

Theorem 2.6. *Given a regular language represented by a DFA or NFA, we can decide in linear time whether the language has polynomial or exponential growth.*

Proof. See, for example, [9]. □

Now let us change focus to sets of integers. Given a subset $S \subseteq \mathbb{N}$ we define

$$\pi_S(x) = \#\{n \leq x \colon n \in S\}. \tag{2.1}$$

If there exists an integer $d \geq 0$ such that $\pi_S(x) = O((\log x)^d)$, then we say that S is *sparse*. Otherwise we say S is *non-sparse*.

Then the corollary below follows immediately from the above results.

Corollary 2.7. *Let $k \geq 2$ be an integer and let S be a k-automatic subset of $\mathbb{N}$. Then S is non-sparse iff there exists a real number $\alpha > 0$ such that $\pi_S(x) = \Omega(x^\alpha)$.*

Given sets S, T of real numbers, we let $S + T$ denote the set

$$\{s + t \ : \ s \in S, t \in T\}.$$

Furthermore, we let $S^j = \overbrace{S + S + \cdots + S}^{j}$; this is called the j-fold sum of S. We let $S^{\leq j} = \bigcup_{1 \leq i \leq j} S^i$. Note that $S^{\leq j}$ and S^j denote, respectively, the set of numbers that can be written as a sum of at most j elements of S and those that can be written as a sum of exactly j elements of S. Finally, if S is a set of real numbers and α is a real number, then $\alpha S = \{\alpha x \ : \ x \in S\}$.

3. Sums of Cantor sets

In this section, we quickly recall the basic notions we will make use of concerning Cantor sets. Specifically, we will be dealing with central Cantor sets, which we now define. Let $(r_k)_{k\geq 1}$ be a sequence of real numbers in the half-open interval $(0, \frac{1}{2}]$. Given real numbers $\alpha < \beta$, we define a collection of closed intervals $\{C_w : w \in \{0,1\}^*\}$, where each $C_w \subseteq [\alpha, \beta]$, inductively as follows. We begin with $C_\epsilon = [\alpha, \beta]$. Having defined C_w for all binary words of length at most n, given a word w of length $n+1$, we write $w = w'a$ with $|w'| = n$ and $a \in \{0,1\}$. If $a = 0$, we define C_w to be the closed interval uniquely defined by having the same left endpoint as $C_{w'}$ and satisfying $|C_w|/|C_{w'}| = r_{n+1}$. If $a = 1$, we define C_w to be the closed interval uniquely defined by having the same right endpoint as $C_{w'}$ and satisfying $|C_w|/|C_{w'}| = r_{n+1}$. We then take C_n to be the union of the C_w as w ranges over words of length n. It is straightforward to see that

$$C_0 \supseteq C_1 \supseteq C_2 \supseteq \cdots,$$

and the intersection of these sets is called the *central Cantor set* associated with the ratios r_k and initial interval $[\alpha, \beta]$. The associated real numbers r_k are called the associated *ratios of dissection*, and in the case when there is a fixed r such that $r_k = r$ for every $k \geq 1$, we simply call r the ratio of dissection. A key example is the classical "middle thirds" Cantor set, which is the central Cantor set with ratio of dissection $\frac{1}{3}$ and initial interval $[0, 1]$.

Let $k \geq 2$ be a natural number and let $u, y, z \in \Sigma_k^*$ with $|y| = |z|$ and $y \neq z$. In particular, y and z are non-empty. We define $C(u; y, z)$ to be the collection of real numbers whose base-k expansion is of the form $0.uw_1w_2w_3\cdots$ with each $w_i \in \{y, z\}$. For example, when $k = 3$, u is the empty word, $y = 0$, and $z = 2$, $C(u; y, z)$ is the usual Cantor set. A key lemma used in our considerations rests on a result of Cabrelli, the second-named author, and Molter [7], which says that a set formed by taking the sum of N elements from a Cantor set with a fixed ratio of dissection is equal to an interval when N is sufficiently large. We use this result to prove the following lemma.

Lemma 3.1. *Let $k \geq 2$ and $t \geq 1$ be natural numbers and let $u, y, z \in \Sigma_k^*$ with $|y| = |z|$ and $y \neq z$. Suppose that $|u| = L$ and $|y| = |z| = s$. Then every real number $\gamma \in [k^{L+s+1}, k^{L+s+1+t}]$ can be expressed as a sum of at most $k^{2L+2s+t+1}$ elements from $C(u; y, z)$.*

Proof. Let $s = |y| = |z|$ and write $y = y_1 \cdots y_s$, $z = z_1 \cdots z_s$, and $u = u_1 \cdots u_L$. Define

$$Y = \sum_{j=1}^{s} y_j k^{-j},$$

$$Z = \sum_{j=1}^{s} z_j k^{-j},$$

$$U = \sum_{j=1}^{L} u_j k^{-j}.$$

We may assume without loss of generality that $Y < Z$. Consider the compact set $C = C(\epsilon; y, z)$, the numbers whose base-k expansion is of the form $0.x_1x_2x_3\cdots$ where $x_i \in \{y, z\}$. The two contractions $S_1(x) = k^{-s}x + Y$ and $S_2(x) = k^{-s}x + Z$ clearly map C into C; hence C contains $S_1(C) \cup S_2(C)$. We claim that this containment is in fact an equality. To see this, let x be a real number with base-k expansion $0.x_1x_2x_3\cdots$ with $x_i \in \{y, z\}$. Then x is mapped to $0.yx_1x_2\cdots$ under S_1 and to $0.zx_1x_2\cdots$ under S_2. In particular, $x = S_1(0.x_2x_3\cdots)$ if $x_1 = y$ and $x = S_2(0.x_2x_3\cdots)$ if $x_1 = z$.

Next, consider C', the set obtained by beginning with the non-trivial interval $[\alpha, \beta]$ where $\alpha = (1 - k^{-s})^{-1}Y$ and $\beta = (1 - k^{-s})^{-1}Z$ and forming the central Cantor set with ratio of dissection k^{-s}.

Then C' also has the property that $C' = S_1(C') \cup S_2(C')$. Indeed, the set C'_n that arises at level n in the Cantor set construction is the union of the images of $[\alpha, \beta]$ under the n-fold compositions $S_{j_1} \circ \cdots \circ S_{j_n}$, where $j_i \in \{1, 2\}$ for $i = 1, \ldots, n$. Then C' is simply the intersection of the C'_n for $n \geq 1$.

Since there is a unique non-empty compact set with the above invariance property under the two contractions S_1 and S_2, we must have $C = C'$. Thus C has a central Cantor set construction with ratio of dissection k^{-s}. It now follows from [7, Proposition 2.2] that the m-fold sum C^m equals the interval $[m\alpha, m\beta]$ whenever $m \geq k^s - 1$.

The set $C(u; y, z)$ is equal to $\sum_{j=1}^{L} u_j k^{-j} + k^{-L}C := U + k^{-L}C$. Observe that if $C^m = [c, d]$, then $(k^{-L}C)^m = [k^{-L}c, k^{-L}d]$ and the m-fold sum of $U + k^{-L}C$ is simply the interval $mU + [k^{-L}c, k^{-L}d]$. Thus for all $m \geq k^s - 1$, $C(u; y, z)^m$ contains the non-trivial interval mI where $I = [U + k^{-L}\alpha, U + k^{-L}\beta]$. The intervals mI and $(m + 1)I$ overlap whenever

$$(m + 1)(U + k^{-L}\alpha) \leq m(U + k^{-L}\beta),$$

which occurs precisely when $m \geq (k^L U + \alpha)(\beta - \alpha)^{-1}$. Since $\beta - \alpha \geq 1/k^s$ and $U, \alpha \leq 1$, we see that for $m \geq k^{L+s} + k^s$, the intervals mI and $(m + 1)I$ overlap. Thus

$$\bigcup_{m \geq k^{L+s}+k^s} mI \supseteq [k^{L+s+1}, \infty).$$

Consequently, we have that the interval $[k^{L+s+1}, k^{L+s+1+t}]$ is contained in the union of the m-fold sums of $C(u; y, z)$ with $m = k^{L+s} + k^s, \ldots, N$ whenever N is such that $N(U + k^{-L}\beta) \geq k^{L+s+t+1}$. Since $U + k^{-L}\beta \geq k^{-L-s}$ we see that we can take $N = k^{2L+2s+t+1}$. This proves that every number in $[k^{L+s+1}, k^{L+s+1+t} - 1]$ can be expressed as a sum of at most N elements from $C(u; y, z)$. □

4. The first main result

In this section we prove the following theorem.

Theorem 4.1. *Let $k \geq 2$ be a natural number and let S be a non-sparse k-automatic subset of $\mathbb{N}$ with $\gcd(S) = 1$. Then there exist effectively computable natural numbers $N = N(S)$ and $M = M(S)$ such that every natural number $n \geq M$ can be expressed as a sum of at most N elements from S. Moreover, if the minimal DFA accepting S has m states, then $N \leq 5k^{16m+3}$ and $M \leq 3k^{16m+5}$.*

Remark 4.2. We note that the non-sparse and gcd hypotheses on S are, in fact, necessary to obtain the conclusion of the statement of the theorem.

If $\gcd(S) = g > 1$, then every sum of elements of S is divisible by g.

On the other hand, if S is a sparse k-automatic set, then $\pi_S(x) = O((\log x)^d)$ for some $d \geq 0$. In particular, there is some $C > 0$ such that for all $x \geq 2$ there are at most $C(\log x)^d$ elements of S that are $< x$. Thus there are at most $C^i(\log x)^{di}$ elements of S smaller than x that can be written as the sum of i elements of S. Hence there are at most $\sum_{0 \leq i \leq I} C^i(\log x)^{di}$ elements of S smaller than x that can be written as the sum of at most I elements of S. But this is $O((\log x)^{dI+1})$, which for large x is smaller than x.

This remark combined with Theorem 4.1 easily gives Theorem 1.1.

Remark 4.3. The bounds in Theorem 4.1 are close to optimal. If one considers the set S of all natural numbers whose base-k expansion has j digits, for $j \geq 0$ and $j \equiv -1 \pmod{m}$, then the minimal DFA accepting S has size m. On the other hand, every element of S has size at least k^{m-2}. So for each natural number $d \geq 1$ the interval $[1, k^{md-2} - 1] \cap S$ has size at most $k^{m(d-1)-1} - 1$. Thus $k^{md-2} - 1$ cannot be expressed as a sum of fewer than k^{m-2} elements of S for $m \geq 2$.

Before we prove Theorem 4.1, we need some auxiliary results. We recall that a subset T of the natural numbers is *c-syndetic* for a natural number c if $n \in T$ implies that there exists $i \in \{1, \ldots, c\}$ such that $n + i \in T$. If T is c-syndetic for some c, we say that T is *syndetic*.

Proposition 4.4. *Let $k \geq 2$ be a natural number and let S be a non-sparse k-automatic subset of the natural numbers whose minimal accepting DFA has m states. If T is the set of all numbers that can be written as a sum of at most k^{11m+1} elements of S, then for each $M > k^{7m+1}$ there exists $n \in T$ such that $|M - n| < k^{12m+1}$. In particular, T is $(2k^{12m+1})$-syndetic.*

Proof. Since S is non-sparse, by Lemma 2.4 we have that there exist words $u, y, z, v \in \Sigma_k^*$ with $y \neq z$ and $|u|, |v| \leq m$, $|y| = |z| \leq 3m$ such that $\mathcal{L}(S)$ contains $u\{y, z\}^*v$. Let $L = |u|$ and $s = |y| = |z|$. By Lemma 3.1, taking $t = s$, each $\alpha \in [k^{L+s+1}, k^{L+2s+1}]$ can be expressed as a sum of at most $k^{2L+3s+1} \leq k^{11m+1}$ elements from $C(u; y, z)$.

Now let $0 \leq \alpha < \beta < 1$ be real numbers. Suppose that M is a natural number with base-k expansion $x_0 x_1 \cdots x_d$ (and $x_0 \neq 0$) with $d \geq \max(L + 2s + 1, K + 2L + s + 2)$. We let x denote the k-adic rational number with base-k expansion $0.x_0 x_1 \cdots x_d$. Then for $j \in \{0, 1, \ldots, s-1\}$, the number $k^{L+s+2+j}x$ has base-k expansion

$$x_0 x_1 \cdots x_{L+s+j+1}.x_{L+s+j+2} \cdots x_d \in [k^{L+s+1}, k^{L+2s+1}],$$

and so by Lemma 3.1 there exist $r \leq k^{2L+3s+1}$ and $y_1, \ldots, y_r \in C(u; y, z)$ such that $y_1 + \cdots + y_r = k^{L+s+2+j}x$.

Let ℓ be a positive integer and let $C_\ell(u, v; y, z)$ denote the set of k-adic rationals whose base-k expansions are of the form $0.uw_1 w_2 \cdots w_\ell v$ with $w_1, \ldots, w_\ell \in \{y, z\}$ and let K denote the length of v. Observe that given $\epsilon > 0$ we have that there is a natural number N such that whenever $x \in C(u; y, z)$ and $\ell > N$ there exists $x' \in C_\ell(u, v; y, z)$ such that $|x - x'| < k^{-\ell s - L}$. In particular, there exist $y_{1,\ell}, y_{2,\ell}, \ldots, y_{r,\ell} \in C_\ell(u, v; y, z)$ such that $|y_{i,\ell} - y_i| < k^{-\ell s - L}$ for $i = 1, \ldots, r$.

Thus

$$|y_{1,\ell} + \cdots + y_{r,\ell} - k^{L+s+2+j}x| < rk^{-\ell s - L} \leq k^{2L+3s+1}k^{-\ell s - L} = k^{L+(3-\ell)s+1}.$$

Observe that $k^{L+\ell s+K}y_{i,\ell} \in S$ for $i = 1, \ldots, r$ and so $k^{L+\ell s+K}y_{1,\ell}+\cdots+k^{L+\ell s+K}y_{r,\ell}$ is a sum of at most $k^{2L+3s+1}$ elements of S. By construction it is at a distance of at most $k^{L+\ell s+K}k^{L+(3-\ell)s+1} = k^{2L+3s+K+1}$ from $k^{(\ell+1)s+2L+K+2+j}x$. Since j can take any value in $\{0, 1, \ldots, s-1\}$ and since $d > K + 2L + s + 2$, we see that we can find an element in $S^{\leq r}$ that is at a distance of at most $k^{2L+3s+K+1}$ from M. Finally, since $L + 2s + 1, K + 2L + s + 2 \leq 7m + 1$ and $2L + 3s + K + 1 \leq 12m + 1$, we obtain the desired result. □

Before proving Theorem 4.1 we need two final results about automatic sets.

Lemma 4.5. *Let $k \geq 2$, and suppose $S \subseteq \mathbb{N}$ is a k-automatic set whose minimal accepting DFA has m states. If $\gcd(S) = 1$, then there exist distinct integers $s_1, s_2, \ldots, s_\ell \in S$, all less than k^{2m+2}, such that $\gcd(s_1, s_2, \ldots, s_\ell) = 1$.*

Proof. If $1 \in S$, there is nothing to prove, so we may assume that $1 \notin S$. Let N denote the smallest natural number such that $\gcd(S \cap [1, N+1]) = 1$ and let $d = \gcd(S \cap [1, N])$. In particular, $\gcd(d, N+1) = 1$. By assumption, $d > 1$. We claim that $N \leq k^{2m+2}$. We write $d = k_0 d_0$, where $\gcd(d_0, k) = 1$ and with k_0 dividing a power of k.

We first consider the case when $k_0 > 1$. Let $a \in \{0, 1, \ldots, k-1\}$ be such that $N + 1 \equiv a \pmod{k}$. Then $\gcd(a, k_0) = 1$ since if this is not the case then there is some prime p that divides both a, d, and k, and so p would divide $N+1$ and d, which is a contradiction. Then notice that $S_a := \{n \geq 0 \colon kn + a \in S\}$ contains $(N + 1 - a)/k$ and contains no natural number smaller than $(N + 1 - a)/k$, since if $kn + a \in S$ for some $n < (N + 1 - a)/k$, then $d|(kn + a)$ and so $k_0|(kn + a)$. But this is impossible, because if p is a prime that divides k_0 (and consequently k), then it must divide a, which we have shown cannot occur. Notice that S_a must have a minimal accepting DFA with at most m states. But it is straightforward to see that a non-empty set whose minimal accepting DFA has at most m states must contain an element of size at most k^m and so $N + 1 < k^{m+1} + k$.

Next consider the case when $k_0 = 1$, so $\gcd(d, k) = 1$. We let $t_s \cdots t_0$ denote the base-k expansion of $N+1$. We claim that $s \leq 2m$. To see this, suppose that $s > 2m$ and let $T_i := \{n \geq 0 \colon k^{i+1}n + [t_i \cdots t_0]_k \in S\}$ for $i = 0, \ldots, m$. Then since the minimal DFA accepting S has m states we see there exist $i, j \leq m$ with $i < j$ such that $T_i = T_j$. Also, since each T_ℓ has a minimal accepting DFA with at most m states and each T_ℓ is non-empty, we have that there is some least element $r_\ell \in T_\ell$ with $r_\ell < [t_s \cdots t_{\ell+1}]_k \in T_\ell$. Observe that $r'_\ell := k^{\ell+1}r_\ell + [t_\ell \cdots t_0]_k < N + 1$ and so d divides r'_ℓ. Moreover, for all $r < [t_s \cdots t_{\ell+1}]_k$ with $r \in T_\ell$ we have $k^{\ell+1}r + [t_\ell \cdots t_0]_k \equiv 0 \pmod{d}$. Thus since k and d are relatively prime, we see that $T_\ell \cap [0, [t_s \cdots t_{\ell+1}]_k - 1]$ is non-empty and contained in a single arithmetic progression of difference d, but $[t_s \cdots t_{\ell+1}]_k$ is not in this arithmetic progression.

But now we have that $T_i = T_j$ with $i < j$ and so $T_j \cap [0, [t_s \cdots t_{i+1}]_k - 1]$ is contained in a single arithmetic progression mod d. On the other hand, $T_j \cap [0, [t_s \cdots t_{j+1}]_k - 1]$ is non-empty and contained in a single arithmetic progression mod d, and by the above remarks, $[t_s \cdots t_{j+1}]_k < [t_s \cdots t_{i+1}]_k$ is not in this progression, a contradiction. Thus we see that $s \leq 2m$ and so $N < k^{2m+2}$. □

Lemma 4.6. *Let $k \geq 2$, let m and c be natural numbers, and let $S \subseteq \mathbb{N}$ be a k-automatic set with $\gcd(S) = 1$ whose minimal accepting DFA has m states. If U is the set of elements that can be expressed as a sum of at most $2ck^{4m+2}$ elements of S, then there is some $N \leq ck^{4m+4}$ such that U contains $\{N, N+1, \ldots, N+c\}$.*

Proof. From Lemma 4.5 we know there exist $s_1, s_2, \ldots, s_\ell \in S$ with $s_1 < \cdots < s_\ell \leq k^{2m+2}$ such that $\gcd(s_1, \ldots, s_\ell) = 1$.

It follows from a result of Borosh and Treybig [4, Theorem 1] that there exist integers $a_1, \ldots, a_\ell \in \mathbb{Z}$ with $|a_i| \leq k^{2m+2}$ such that $\sum a_i s_i = 1$.

Now let $t = ck^{2m+2}$ and consider the number $N := ts_1 + \cdots + ts_\ell$. For each $i = 1, \ldots, c$ we have that $N+i = (t+ia_1)s_1 + \cdots + (t+ia_\ell)s_\ell$ is a non-negative integer linear combination of $s_1, \ldots, s_\ell$ and $|t + ia_j| \leq 2ck^{2m+2}$ for $j \in \{1, \ldots, \ell\}$. Thus we see that if U is the set of integers that can be expressed as at most $2ck^{2m+2}\ell$ elements of S, then U contains $\{N, N+1, \ldots, N+c\}$, where $N = ts_1 + \cdots + ts_\ell \leq ck^{2m+2}\ell$. Since $\ell \leq k^{2m+2}$, we obtain the desired result. $\square$

We are now ready for the proof of our first main result.

Proof of Theorem 4.1. Let m be the size of the minimal accepting DFA for S. By Proposition 4.4 if T is the set of elements that can be expressed as the sum of at most k^{11m+1} elements of S, then T is $2k^{12m+1}$-syndetic. Let $c = 2k^{12m+1}$. By assumption $\gcd(S) = 1$, and so by Lemma 4.6 there is some $N_1 \leq 2ck^{4m+2} = 4k^{16m+3}$ and some natural number $M_1 \leq ck^{4m+4} \leq 2k^{16m+5}$ such that each element from $\{M_1, M_1+1, \ldots, M_1+c\}$ can be expressed as a sum of at most N_1 elements of $\{s_1, \ldots, s_d\} \subseteq S$. Then let M_0 denote the smallest natural number in T. Since $T \supseteq S$ and the minimal DFA for S has size at most m, we see that $M_0 \leq k^m$.

We claim that every natural number that is greater than $M := M_0 + M_1 \leq 3k^{16m+5}$ can be expressed as a sum of at most $N := k^{11m+1} + N_1 \leq 5k^{16m+3}$ elements of S. To see this, suppose, in order to get a contradiction, that this is false. Then there is some smallest natural number $n > M$ that cannot be expressed as a sum of at most N elements of S. Observe that $n - M_1 > M_0$; since T is syndetic and $M_0 \in T$, there is some $t \in T$ with $t \leq n - M_1 < t + c$. Thus $n = t + M_1 + j$ for some $j \in \{0, 1, \ldots, c-1\}$. Since $M_1 + j$ is a sum of at most N_1 elements of S and t is the sum of at most k^{11m+1} elements of S, we see that n is the sum of at most N elements of S, contradicting our assumption that n has no such representation. The result follows. $\square$

5. An algorithm

In this section, we prove Theorem 1.2, giving an algorithm to find the smallest number j (if it exists) such that S is an asymptotic additive basis (resp., additive basis) of order j for the natural numbers, where S is a k-automatic set of natural numbers. We use the fact that there is an algorithm for deciding the truth of first-order propositions (involving $+$ and $\leq$) about automatic sequences [1,6,8].

Proof of Theorem 1.2. From Theorem 4.1 and Remark 4.2, we know that S forms an asymptotic additive basis of order j, for some j, if and only if S is non-sparse and has gcd 1. This sparsity criterion can be tested using Lemma 2.1. The condition $\gcd(S) = 1$ can be tested as follows: compute the smallest non-zero member m of S, if it exists. Then $\gcd(S)$ must be a divisor of m. For each divisor d of m, form the assertion

$$\forall n \geq 0 \ (n \in S) \implies \exists t \text{ such that } n = dt$$

and check it using the algorithm for first-order predicates mentioned above. (Note that for each invocation d is actually a constant, so that td actually is shorthand for $\overbrace{t+t+\cdots+t}^{d}$, which uses addition and not multiplication.) The largest such d equals $\gcd(S)$.

Once S passes these two tests, we can test if S is an asymptotic additive basis of order j by writing and checking the predicate

$$\exists M\ \forall n \geq M\ \ \exists x_1, x_2, \ldots, x_j \text{ such that } x_1, x_2, \ldots, x_j \in S\ \wedge\ n = x_1 + x_2 + \cdots + x_j, \tag{5.2}$$

which says every sufficiently large integer is the sum of j elements of S. We do this for $j = 1, 2, 3, \ldots$ until the smallest such j is found. This algorithm is guaranteed to terminate in light of Theorem 4.1.

Finally, once j is known, the optimal M in (5.2) can be determined as follows by writing the predicate in (5.2) together with the assertion that M is the smallest such integer. Using the decision procedure mentioned above, one can effectively create a DFA accepting $(M)_k$, which can then be read off from the transitions of the DFA.

To test if S is an additive basis of order j, we need, in addition to the non-sparseness of S and $\gcd(S) = 1$, the condition $1 \in S$, which is easily checked. If S passes these tests, we then write and check the predicate

$$\forall n \geq 0\ \ \exists x_1, x_2, \ldots, x_j \text{ such that } x_1, x_2, \ldots, x_j \in S\ \wedge\ n = x_1 + x_2 + \cdots + x_j,$$

which says every integer is the sum of j elements of S. We do this for $j = 1, 2, 3, \ldots$ until the least such j is found. □

Remark 5.1. The same kind of idea can be used to test if every element of $\mathbb{N}$ (or every sufficiently large element) is the sum of j *distinct* elements of a k-automatic set S. For example, if $j = 3$, we would have to add the additional condition that

$$x_1 \neq x_2\ \wedge\ x_1 \neq x_3\ \wedge\ x_2 \neq x_3.$$

We can also test if every element is *uniquely* representable as a sum of j elements of S. Similarly, we can count the number $f(n)$ of representations of n as a sum of j elements of S. It follows from [8] that, for k-automatic sets S, the function $f(n)$ is k-regular and one can give an explicit representation for it.

6. Examples

In this section, we give some examples that illustrate the power of the algorithm provided in the preceding section.

Example 6.1. Let S be the 3-automatic set of Cantor numbers

$$\mathcal{C} = \{0, 2, 6, 8, 18, 20, 24, 26, 54, 56, 60, 62, 72, 74, 78, 80, 162, \ldots\},$$

that is, those natural numbers (including 0) whose base-3 expansions consist of only the digits 0 and 2. Then every even number is the sum of exactly two elements of $\mathcal{C}$. To see this, consider an even natural number N. Write $N/2 = x + y$, choosing the base-3 expansions of x and y digit-by-digit as follows:

(a) if the digit of $N/2$ is 2, choose 1 for the corresponding digit in both x and y;
(b) if the digit of $N/2$ is 1, choose 1 for the corresponding digit in x and 0 for the corresponding digit in y;
(c) if the digit of $N/2$ is 0, choose 0 for the corresponding digit in both x and y.

Then $N = 2x + 2y$ gives the desired representation.

Example 6.2. Let S be the 2-automatic set of "evil" numbers

$$\mathcal{E} = \{0, 3, 5, 6, 9, 10, 12, 15, 17, 18, 20, 23, 24, 27, 29, 30, 33, 34, 36, 39, \ldots\},$$

that is, those natural numbers (including 0) for which the sum of the binary digits is even (see, e.g., [3, p. 431]). Then every integer other than $\{1, 2, 4, 7\}$ is the sum of three elements of $\mathcal{E}$. In fact, every integer except $\{2, 4\} \cup \{2 \cdot 4^i - 1 \ : \ i \geq 1\}$ is the sum of two elements of $\mathcal{E}$.

Example 6.3. Let S be the 2-automatic set

$$\begin{aligned}\mathcal{R} &= \{n \ : r(n) = -1\} \\ &= \{3, 6, 11, 12, 13, 15, 19, 22, 24, 25, 26, 30, 35, 38, 43, 44, 45, 47, \ldots\},\end{aligned}$$

where $r(n)$ is the Golay-Rudin-Shapiro function [11, 12, 26, 27]. Then every integer except $\{0, 1, 2, 3, 4, 5, 7, 8, 10, 11, 13, 20\}$ is the sum of two elements of $\mathcal{R}$.

Example 6.4. Let S be the 4-automatic set

$$\mathcal{D} = \{0, 1, 4, 5, 16, 17, 20, 21, 64, 65, 68, 69, 80, 81, 84, 85, \ldots\}$$

of integers representable in base 4 using only the digits 0 and 1. See, for example, [5, 20]. Then every natural number is representable as the sum of three elements of $\mathcal{D}$. In fact, even more is true: every natural number is uniquely representable as the sum of one element chosen from $\mathcal{D}$ and one element chosen from $2\mathcal{D}$.

All these examples can be proved "automatically" by the `Walnut` theorem-proving software [22]. On a laptop each proof is completed within 21 milliseconds.

7. Concluding remarks

It is natural to ask if our results can be extended to k-context-free sets [17, 21]. In this more general setting, however, Theorem 1.1 no longer holds. The following example was shown to the third author by a participant of the Knuth 80 conference held in January 2018 in Piteå, Sweden.

Example 7.1. Let $k = 2$ and consider the set

$$S = \{n \ : \ (n)_2 = 10^n x, \text{ where } |x| = n \geq 0\} \cup \{0\}.$$

It is easy to see that this set is 2-context-free. Furthermore, we have $\pi_S(x) = \Theta(x^{1/2})$.

However, to represent numbers of the form $2^n(2^n - 1)$ with elements of S we need at least $n/2$ summands, because adding together elements of S in decreasing order introduces at most two additional 1's in the high order bits with each additional summand. One 1 bit can derive from the block at the beginning, and the other can derive from a carry from the least significant digits.

So S cannot have an asymptotic basis of any finite order, even though it has the right density.

References

[1] Jean-Paul Allouche, Narad Rampersad, and Jeffrey Shallit, *Periodicity, repetitions, and orbits of an automatic sequence*, Theoret. Comput. Sci. **410** (2009), no. 30-32, 2795–2803, DOI 10.1016/j.tcs.2009.02.006. MR2543333

[2] Jean-Paul Allouche and Jeffrey Shallit, *Automatic sequences*, Theory, applications, generalizations, Cambridge University Press, Cambridge, 2003. MR1997038

[3] Elwyn R. Berlekamp, John H. Conway, and Richard K. Guy, *Winning ways for your mathematical plays. Vol. 2*, Games in particular, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1982. MR654502

[4] I. Borosh and L. B. Treybig, *Bounds on positive integral solutions of linear Diophantine equations*, Proc. Amer. Math. Soc. **55** (1976), no. 2, 299–304, DOI 10.2307/2041711. MR0396605

[5] N. G. de Bruijn, *Some direct decompositions of the set of integers*, Math. Comp. **18** (1964), 537–546, DOI 10.2307/2002940. MR0167447

[6] Véronique Bruyère, Georges Hansel, Christian Michaux, and Roger Villemaire, *Logic and p-recognizable sets of integers*, Bull. Belg. Math. Soc. Simon Stevin **1** (1994), no. 2, 191–238. Journées Montoises (Mons, 1992). MR1318968

[7] Carlos A. Cabrelli, Kathryn E. Hare, and Ursula M. Molter, *Sums of Cantor sets*, Ergodic Theory Dynam. Systems **17** (1997), no. 6, 1299–1313, DOI 10.1017/S0143385797097678. MR1488319

[8] Émilie Charlier, Narad Rampersad, and Jeffrey Shallit, *Enumeration and decidable properties of automatic sequences*, Internat. J. Found. Comput. Sci. **23** (2012), no. 5, 1035–1066, DOI 10.1142/S0129054112400448. MR2983367

[9] Paweł Gawrychowski, Dalia Krieger, Narad Rampersad, and Jeffrey Shallit, *Finding the growth rate of a regular or context-free language in polynomial time*, Internat. J. Found. Comput. Sci. **21** (2010), no. 4, 597–618, DOI 10.1142/S0129054110007441. MR2678190

[10] Seymour Ginsburg and Edwin H. Spanier, *Bounded regular sets*, Proc. Amer. Math. Soc. **17** (1966), 1043–1049, DOI 10.2307/2036087. MR0201310

[11] M. J. E. Golay, *Multi-slit spectrometry*, J. Optical Soc. America **39** (1949), 437–444.

[12] M. J. E. Golay, *Static multislit spectrometry and its application to the panoramic display of infrared spectra*, J. Optical Soc. America **41** (1951), 468–472.

[13] John E. Hopcroft and Jeffrey D. Ullman, *Introduction to automata theory, languages, and computation*, Addison-Wesley Series in Computer Science, Addison-Wesley Publishing Co., Reading, Mass., 1979. MR645539

[14] Oscar H. Ibarra and B. Ravikumar, *On sparseness, ambiguity and other decision problems for acceptors and transducers*, STACS 86 (Orsay, 1986), Lecture Notes in Comput. Sci., vol. 210, Springer, Berlin, 1986, pp. 171–179, DOI 10.1007/3-540-16078-7_74. MR827734

[15] D. M. Kane, C. Sanna, and J. Shallit, *Waring's theorem for binary powers*, preprint, available at `https://arxiv.org/abs/1801.04483`. Submitted 2018.

[16] G. Kreisel, D. Lacombe, and J. R. Shoenfield, *Partial recursive functionals and effective operations*, Constructivity in mathematics: Proceedings of the colloquium held at Amsterdam, 1957 (edited by A. Heyting), Studies in Logic and the Foundations of Mathematics, North-Holland Publishing Co., Amsterdam, 1957, pp. 290–297. MR0108443

[17] Marion Le Gonidec, *On the complexity of a family of k-context-free sequences*, Theoret. Comput. Sci. **414** (2012), 47–54, DOI 10.1016/j.tcs.2011.09.022. MR2896582

[18] M. Lothaire, *Combinatorics on words*, with a foreword by Roger Lyndon and a preface by Dominique Perrin; corrected reprint of the 1983 original, with a new preface by Perrin, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1997. MR1475463

[19] P. Madhusudan, D. Nowotka, A. Rajasekaran, and J. Shallit. *Lagrange's theorem for binary squares*, to appear, 43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018), available at `https://arxiv.org/abs/1710.04247`. Submitted 2017.

[20] Leo Moser, *An application of generating series*, Math. Mag. **35** (1962), no. 1, 37–38. MR1571147

[21] Yossi Moshe, *On some questions regarding k-regular and k-context-free sequences*, Theoret. Comput. Sci. **400** (2008), no. 1-3, 62–69, DOI 10.1016/j.tcs.2008.02.018. MR2424342

[22] H. Mousavi. *Automatic theorem proving in Walnut*, preprint available at `https://arxiv.org/abs/1603.06017`, 2016.

[23] Melvyn B. Nathanson, *Additive number theory*, The classical bases, Graduate Texts in Mathematics, vol. 164, Springer-Verlag, New York, 1996. MR1395371

[24] A. Rajasekaran, *Using automata theory to solve problems in additive number theory*, Master's thesis, School of Computer Science, University of Waterloo, 2018.

[25] A. Rajasekaran, T. Smith, and J. Shallit, Sums of palindromes: an approach via automata, in R. Niedermeier and B. Vallée, editors, *35th Symposium on Theoretical Aspects of Computer Science (STACS 2018)*, Leibniz International Proceedings in Informatics, pages 54:1–54:12. Schloss Dagstuhl — Leibniz-Zentrum für Informatik, 2018.

[26] Walter Rudin, *Some theorems on Fourier coefficients*, Proc. Amer. Math. Soc. **10** (1959), 855–859, DOI 10.2307/2033608. MR0116184

[27] H. S. Shapiro, *Extremal problems for polynomials and power series*, Master's thesis, MIT, 1952.

[28] Andrew Szilard, Sheng Yu, Kaizhong Zhang, and Jeffrey Shallit, *Characterizing regular languages with polynomial densities*, Mathematical foundations of computer science 1992 (Prague, 1992), Lecture Notes in Comput. Sci., vol. 629, Springer, Berlin, 1992, pp. 494–503, DOI 10.1007/3-540-55808-X_48. MR1255157

[29] V. I. Trofimov, *Growth functions of some classes of languages*; Russian, with English summary transl., Cybernetics **6** (1981), i, 9–12, 149. MR689421

[30] R. C. Vaughan and T. D. Wooley, *On Waring's problem: some refinements*, Proc. London Math. Soc. (3) **63** (1991), no. 1, 35–68, DOI 10.1112/plms/s3-63.1.35. MR1105718

[31] Bin Wei and Trevor D. Wooley, *On sums of powers of almost equal primes*, Proc. Lond. Math. Soc. (3) **111** (2015), no. 5, 1130–1162, DOI 10.1112/plms/pdv048. MR3477231

[32] Trevor D. Wooley, *Large improvements in Waring's problem*, Ann. of Math. (2) **135** (1992), no. 1, 131–164, DOI 10.2307/2946566. MR1147960

[33] Dan Zwillinger, *A Goldbach conjecture using twin primes*, Math. Comp. **33** (1979), no. 147, 1071, DOI 10.2307/2006081. MR528060

Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Email address: `jpbell@uwaterloo.ca`

Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Email address: `kehare@uwaterloo.ca`

School of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Email address: `shallit@uwaterloo.ca`