Quantitative estimates on the size of an intersection of sparse automatic sets

Seda Albayrak
(Joint work with Jason Bell)

Department of Maths&Stats, University of Calgary
Calgary, AB gulizar.albayrak@ucalgary.ca
https://sites.google.com/view/gsedaa

Outline

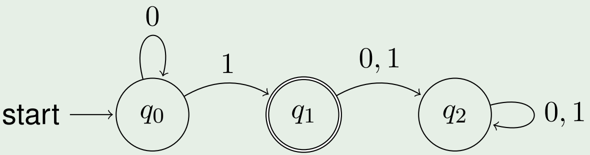
- Breakdown of the title
  - Automatic sets
  - Sparse automatic sets
- Cobham's theorem
- Main result
- Extension
- Conjecture

What is an automatic set?

Definition

Let $k \geq 2$ be a natural number. A subset S of $\mathbb{N}$ is k*-automatic* if there is a finite-state automaton with input alphabet $\Sigma_k = \{0, 1, \ldots, k-1\}$ with the property that the words over the alphabet Σ_k which are accepted by the automaton are precisely the words that are base-k expansions of elements of S.

Example : A finite-state automaton accepting the binary expansions of elements in the set of powers of 2



Notes: 1) reading right-to-left 2) expansions have no leading zeros

What is a sparse automatic set?

Definition

If $S \subseteq \mathbb{N}$, $\pi_S(x) := \#\{n \in S : n \leq x\}$.

A Dichotomy:

For a k-automatic subset $S \subseteq \mathbb{N}$, we have

(1) either there exists an integer $c \geq 1$ such that $\pi_S(x) = O\left((\log x)^c\right)$ as $x \to \infty$,

(2) or there is some $\alpha > 0$ such that $\pi_S(x) > x^\alpha$ for x large.

We call a set $S \subseteq \mathbb{N}$ *sparse k-automatic* if (1) holds.
Otherwise, we call it *non-sparse*.
Question: Where does this dichotomy come from?

Sparse languages have been extensively studied.

e.g. Trofimov (1982) showed this dichotomy for context-free languages in "Growth functions of some classes of languages":

Theorem: The growth function of an arbitrary CF language is either polynomially bounded from above or exponentially bounded from below.

Given a finite alphabet Σ and a language $\mathcal{L} \subseteq \Sigma^*$ over Σ, we have an associated counting function

$$f_{\mathcal{L}}(n) := \#\{w \in \mathcal{L} \colon \mathrm{length}(w) \leq n\}.$$

A regular language $\mathcal{L}$ is *sparse* if $f_{\mathcal{L}}(n) = O(n^d)$ for some natural number d.

We defined a sparse set by translating this to sets.

Example of a sparse automatic set

The set of powers of p is a sparse automatic set, p prime.

In case anyone is wondering "what would be an example of a non-sparse set (or language)?"...:
Consider the Thue-Morse sequence given by

$$t(n) = \begin{cases} 1 & \text{if } s_2(n) \equiv 1 \mod 2 \\ 0 & \text{if } s_2(n) \equiv 0 \mod 2 \end{cases} \tag{1}$$

where $s_2(n)$ is the sum of the digits in the binary expansion of n.
Let T be the set whose characteristic function

$$\chi_T(n) := \begin{cases} 1 & \text{if } n \in T \\ 0 & \text{if } n \notin T \end{cases} \tag{2}$$

is $t(n)$.
T is a non-sparse 2-automatic set of natural numbers.
$\pi_T(n) \sim \frac{n}{2}$.

We are interested in *sparse* (automatic) sets.
Why are they important? Where do sparse sets arise?
Some examples:

- The zero set of a linearly recurrent sequence over a field of characteristic $p > 0$ is a finite union of arithmetic progressions augmented by a sparse p-automatic set (Derksen, Skolem-Mahler-Lech theorem in positive characteristic, 2006)
- Kedlaya's work on extending Christol's theorem to give a full characterization of the algebraic closure of $\mathbb{F}_p(t)$ works by generalizing the notion of automatic sequences to maps $f : S_p \to \mathbb{F}_q$, where S_p is the set of nonnegative elements of $\mathbb{Z}[p^{-1}]$, and as part of his work, he shows that for the maps that arise, the post-radix point behaviour of the support of f can be described in terms of sparse automatic sequences.

Back to our topic:

- Automatic set ✓
- Sparse automatic set ✓
- Next: intersection of sparse automatic sets

Main result is giving
an estimate on the size of intersection of sparse automatic sets.

i.e. the claim is that this intersection is finite.

Keeping in mind that we will be looking at the intersection of sparse automatic sets, consider the following conjecture (now proved):

Catalan's conjecture (1844)

The only solution in the natural numbers of

$$x^a - y^b = 1$$

for $a, b > 1$, $x, y > 0$ is $x = 3$, $a = 2$, $y = 2$, $b = 3$.

Proved in 2002 by Preda Mihăilescu
Method: theory of cyclotomic fields + theory of linear forms in logarithms + short computer computation
Later, he also proved it purely algebraically which does not require a computer calculation.

Keeping in mind that we will be looking at the intersection of sparse automatic sets, consider the following conjecture (now proved):

Catalan's conjecture (1844)

The only solution in the natural numbers of

$$x^a - y^b = 1$$

for $a, b > 1$, $x, y > 0$ is $x = 3$, $a = 2$, $y = 2$, $b = 3$.

(Proved in 2002 by Preda Mihăilescu)

The (*sparse* 2*-automatic*) set consisting of $2^n + 1$, $n \geq 0$

and

the (*sparse* 3*-automatic set*) set consisting of powers of 3

has finite intersection.

Note that 2 and 3 are multiplicatively independent—none of them can be written as a rational power of the other.

Catalan's conjecture is solved but the following isn't:

Conjecture (Erdős, 1979)

For $n \geq 9$, 2^n is not the sum of distinct powers of 3.

e.g. $2^8 = 3^5 + 3^2 + 3 + 1$.
Erdős: "as far as I can see, there is no method at our disposal to attack this conjecture."
Conjecture: the set of powers of 2 and the set consisting of numbers whose ternary expansions omit 2 has finite intersection.
Note that the set of powers of 2 is a 2-*automatic set*
and
the set consisting of numbers whose ternary expansions omit 2 is a 3-*automatic set*.

Theorem (Cobham, 1969)

Let $k, \ell \geq 2$ be two natural numbers that are multiplicatively independent (i.e. there are no non-trivial integer solutions to $k^a = \ell^b$). If $S \subseteq \mathbb{N}$ is a set that is both k- and ℓ-automatic then S is in fact eventually periodic; i.e. there is some fixed positive integer c such that for sufficiently large $n \in \mathbb{N}$, $n \in S$ implies $n + c \in S$.

What happens if S is sparse k-automatic?
Sparse infinite non-empty sets cannot be eventually periodic.
So S cannot be both k- and ℓ-automatic.

A useful characterization of sparse sets:

Theorem (Ginsburg-Spanier, 1966)

A sparse set is a finite union of sets of the form

$$\{[v_0 w_1^* v_1 w_2^* \cdots v_{s-1} w_s^* v_s]_k\}.$$

Fact:

Let $k \geq 2$ be a natural number and let $S \subseteq \mathbb{N}$ be a non-empty simple sparse k-automatic set. Then there exist $s \geq 0$, $c_0, \ldots, c_s \in \mathbb{Q}$ such that $(k^\ell - 1)c_i \in \mathbb{Z}$ for some $\ell \geq 0$, $c_0 + c_1 + \cdots + c_s \in \mathbb{Z}_{\geq 0}$ and positive integers $\delta_1, \ldots, \delta_s$ such that S is of the form

$$\Big\{c_0 + c_1 k^{\delta_s n_s} + c_2 k^{\delta_s n_s + \delta_{s-1} n_{s-1}} \cdots + c_s k^{\delta_s n_s + \cdots + \delta_1 n_1} : n_1, \ldots, n_s \geq 0\}$$

Theorem (A.-Bell, 2023)

Let k and ℓ be multiplicatively independent natural numbers greater than or equal to 2 (i.e., there are no solutions to the equation $k^a = \ell^b$ with nonzero integers a and b). If X is a sparse k-automatic subset of $\mathbb{N}$ and Y is a sparse ℓ-automatic set of $\mathbb{N}$, then $X \cap Y$ is finite.

We prove this by giving an upper bound for the size of the intersection in terms of data from the automata that accept these sets.

Proposition

Let $N \geq 2$, *let* Σ *be a finite alphabet of size* N, *and let* $\Gamma = (Q, \Sigma, \delta, q_0, F)$ *be a deterministic finite automaton accepting a sparse language* $\mathcal{L}$. *Then* $\mathcal{L}$ *is a finite union (possibly empty) of at most*

$$(|Q| - 1)!(N^{|Q|-1} + N^{|Q|-2} + \cdots + 1)$$

languages of the form

$$\{v_0 w_1^* v_1 w_2^* \cdots v_{s-1} w_s^* v_s\}$$

with $w_1, \ldots, w_s, v_1, \ldots, v_s$ *words in* Σ^* *in which the* w_i *are non-empty but the* v_i *may be empty and with* $|w_1| + \cdots + |w_s| \leq |Q| - 1$ *and* $|v_0| + \cdots + |v_s| \leq N(|Q| - 1)$.

Let k and ℓ be multiplicatively independent positive integers and let X be a sparse k-automatic subset of $\mathbb{N}$ of the form

$$\{[v_0 w_1^* v_1 w_2^* \cdots v_s w_s^* v_{s+1}]_k\}$$

and let Y be sparse ℓ-automatic set of the form

$$\{[u_0 y_1^* u_1 y_2^* \cdots u_t y_t^* u_{t+1}]_\ell\}.$$

Then X is of the form

$$\Big\{c_0 + c_1 k^{\delta_s n_s} + c_2 k^{\delta_s n_s + \delta_{s-1} n_{s-1}} \cdots + c_s k^{\delta_s n_s + \cdots + \delta_1 n_1} : n_1, \ldots, n_s \geq 0\Big\},$$

where $c_0, \ldots, c_s$ are rational numbers.

Similarly, Y is of the form

$$\left\{d_0 + d_1 \ell^{\delta'_t m_t} + d_2 \ell^{\delta'_t m_s + \delta'_{t-1} m_{t-1}} \cdots + d_t \ell^{\delta'_t m_t + \cdots + \delta'_1 m_1} : m_1, \ldots, m_t \geq 0\right\},$$

where $d_0, \ldots, d_t$ are rational numbers.
Then an element in $X \cap Y$ corresponds to a solution to the equation

$$d_0 X_0 + \cdots + d_t X_t - c_0 Y_0 - \cdots - c_s Y_s = 0,$$

where $X_0 = 1, X_1 = \ell^{\delta'_t m_t}, \ldots, X_t = \ell^{\delta'_t m_t + \cdots + \delta'_1 m_1}$ and $Y_0 = 1$, ..., $Y_s = k^{\delta_s n_s + \cdots + \delta_1 n_1}$, with the corresponding element in the intersection given by

$$A := d_0 X_0 + \cdots + d_t X_t = c_0 Y_0 + \cdots + c_s Y_s.$$

S-unit theorem

Recall

For $z_1, \ldots, z_n$ in a field K, the equation $z_1 + \cdots + z_n = 1$ is said to be *non-degenerate* if no non-trivial subsum of the left-hand side is equal to zero; that is, whenever I is a nonempty subset of $\{1, \ldots, n\}$, we have $\sum_{i \in I} z_i \neq 0$.

Theorem (Schlickewei, 1990)

Let K be a field of characteristic zero, let $a_1, \ldots, a_n$ be nonzero elements of K, and let $H \subset (K^*)^n$ be a finitely generated multiplicative subgroup. Then there are only finitely many non-degenerate solutions $(x_1, \ldots, x_n) \in H$ to the equation

$$a_1 x_1 + \cdots + a_n x_n = 1.$$

A quantitative version:

Theorem (Amoroso-Viada, 2009)

Let K be a field of characteristic zero, let $a_1, \ldots, a_n$ be nonzero elements of K, and let Γ be a finitely generated multiplicative subgroup of $(K^*)^n$ of rank $r < \infty$. Then there are at most

$$(8n)^{4n^4(n+r+1)}$$

non-degenerate solutions to the equation

$$a_1 x_1 + \cdots + a_n x_n = 1$$

with $(x_1, \ldots, x_n) \in \Gamma$.

We use Amoroso-Viada's theorem to get:

Proposition

Let k and ℓ be multiplicatively independent positive integers and let W be a sparse k-automatic subset of $\mathbb{N}$ of the form

$$\{[v_0 w_1^* v_1 w_2^* \cdots v_{s-1} w_s^* v_s]_k\}$$

where s is a nonnegative integer and $v_0, v_1, \ldots, v_s$ are possibly empty and $w_1, \ldots, w_s$ are non-empty words in Σ_k^*, and let Z be sparse ℓ-automatic set of the form

$$\{[u_0 y_1^* u_1 y_2^* \cdots u_{t-1} y_t^* u_t]_\ell\}$$

where t is a nonnegative integer and $u_0, u_1, \ldots, u_t$ are possibly empty and $y_1, \ldots, y_t$ are non-empty words in Σ_ℓ^*. Then

$$|W \cap Z| \leq (8(s+t+1))^{10(s+t+2)^5 - (s+t+2)}.$$

Combining gives

Theorem (A.-Bell, 2023)

Let k and ℓ be multiplicatively independent positive integers, and let $\Gamma = (Q, \Sigma_k, \delta, q_0, F)$ and $\Gamma' = (Q', \Sigma_\ell, \delta', q_0', F')$ be DFA accepting sparse regular languages $\mathcal{L} \subseteq (\Sigma_k)^*$ and $\mathcal{L}' \subseteq (\Sigma_\ell)^*$. If $X \subseteq \mathbb{N}$ is the set of natural numbers whose base-k expansions are elements of $\mathcal{L}$ and $Y \subseteq \mathbb{N}$ is the set of natural numbers whose base-ℓ expansions are elements of $\mathcal{L}'$, then

$$|X \cap Y| \leq k^{|Q|} \cdot \ell^{|Q'|} \cdot \left(8(|Q| + |Q'| - 1)\right)^{10(|Q|+|Q'|)^5}.$$

Recall

For every natural number n, there is a word $w = (n)_k \in \{0, 1, \ldots, k-1\}^*$, which is called *the base-k expansion of n*. Similarly, given a non-empty word there is a natural number $n = [w]_k$, which is the natural number whose base-k expansion is w. e.g. $[100001]_2 = 33$, $(10)_2 = 1010$.

Extend this notion of automaticity to subsets of $\mathbb{N}^d$ with $d \geq 1$, working with the input alphabet Σ_k^d. Then, given a d-tuple $(n_1, \ldots, n_d)$ of natural numbers, there exist words $w_1, \ldots, w_d$ of the same length with the additional property that w_i is a base-k expansion of n_i for $i = 1, \ldots, d$. e.g. $[2110, 0020]_3 = (66, 6)$. Then a subset S of $\mathbb{N}^d$ is k-automatic if there is a finite-state machine with input alphabet Σ_k^d that accepts precisely the words $(w_1, \ldots, w_d)$ corresponding to d-tuples of natural numbers in S.

Definition

- If $S \subseteq \mathbb{N}^d$,

$$\pi_S(x) := \#\{(n_1, \ldots, n_d) \in S \colon n_1 + n_2 + \cdots + n_d \leq x\}$$

A Dichotomy:

For a k-automatic subset $S \subseteq \mathbb{N}^d$, we have

(1) either there exists an integer $c \geq 1$ such that $\pi_S(x) = O\left((\log x)^c\right)$ as $x \to \infty$,

(2) or there is some $\alpha > 0$ such that $\pi_S(x) > x^\alpha$ for x large.

Example

The set $\{(3^m, 3^m + 1) : m \in \mathbb{N}\}$.

Theorem (Cobham, 1969)

Let $k, \ell \geq 2$ be two natural numbers that are multiplicatively independent (i.e. there are no non-trivial integer solutions to $k^a = \ell^b$). If $S \subseteq \mathbb{N}$ is a set that is both k- and ℓ-automatic then S is in fact eventually periodic; i.e. there is some fixed positive integer c such that for sufficiently large $n \in \mathbb{N}$, $n \in S$ implies $n + c \in S$.

Theorem (Semenov, 1977)

A subset of $\mathbb{N}^d$ that is both k- and ℓ-automatic, with k and ℓ multiplicatively independent, is automatic with respect to all integer bases.

Theorem (A.-Bell, 2023)

Let $k, \ell \geq 2$ be two natural numbers that are multiplicatively independent. If X is a sparse k-automatic subset of $\mathbb{N}^d$ and Y is a sparse ℓ-automatic subset of $\mathbb{N}^d$, then $X \cap Y$ is finite.

Quantitative version:

Theorem (A.-Bell, 2023)

Let k and ℓ be multiplicatively independent positive integers, let $d \geq 2$, and let $\Gamma = (Q, \Sigma_k^d, \delta, q_0, F)$ and $\Gamma' = (Q', \Sigma_\ell^d, \delta', q_0', F')$ be deterministic finite-state automata accepting sparse regular languages $\mathcal{L} \subseteq (\Sigma_k^d)^*$ and $\mathcal{L}' \subseteq (\Sigma_\ell^d)^*$. If $X \subseteq \mathbb{N}^d$ is the set of d-tuples of natural numbers whose base-k expansions are elements of $\mathcal{L}$ and $Y \subseteq \mathbb{N}^d$ is the set of d-tuples of natural numbers whose base-ℓ expansions are elements of $\mathcal{L}'$, then

$$|X \cap Y| \leq k^{d|Q|} \cdot \ell^{d|Q'|} \cdot \big(8(|Q| + |Q'| - 1)\big)^{10d(|Q|+|Q'|)^5}.$$

Method:

- X is a union of "certain number of" simple sparse sets in $\mathbb{N}^d$
- Y is a union of "certain number of" simple sparse sets in $\mathbb{N}^d$
- Then we look at the intersection of the projections of these pairs of simple sparse sets in $\mathbb{N}^d$
- These are simple sparse sets in $\mathbb{N}$.
- We use S-unit theory to find a bound on the size of these intersections.
- then put everything together to get a bound on $|X \cap Y|$.

For a subset S of $\mathbb{N}$, the *density* of S is the limit

$$\lim_{n\to\infty} \frac{\pi_S(n)}{n}, \text{ if it exists.}$$

e.g. Sparse subsets of $\mathbb{N}$ have zero density.

Conjecture (A.-Bell)

Let k, ℓ be multiplicatively independent positive integers. If X is a sparse k-automatic subset of $\mathbb{N}$ and Y is a zero-density ℓ-automatic subset of $\mathbb{N}$, then $X \cap Y$ is finite.

Recall Erdős' conjecture: The set of powers of 2 (which is a sparse 2-automatic set) and the set consisting of numbers whose ternary expansions omit 2 (which is 3-automatic set with zero density) has finite intersection.

References

S. Albayrak and J. P. Bell, A refinement of Christol's theorem for algebraic power series. *Math. Z.* **300** (2022), no. 3, 2265–2288.

F. Amoroso and E. Viada, Small points on subvarieties of a torus. *Duke Math. J.* **150** (2009), no. 3, 407–442.

A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata. *Math. Systems Theory* **3** (1969), 186–192.

P. Erdős, Some unconventional problems in number theory. *Math. Mag.* **52** (1979), no. 2, 67–70.

S. Ginsburg and E. Spanier, Bounded regular sets. *Proc. Amer. Math. Soc.* **17** (1966), 1043–1049.

Mihăilescu, Preda, Primary cyclotomic units and a proof of Catalan's conjecture. *J. Reine Angew. Math.* **572** (2004), 167–195.

A. L. Semenov, The Presburger nature of predicates that are regular in two number systems. (Russian) *Sibirsk. Mat. Ž.* **18** (1977), no. 2, 403–418, 479.

H. P. Schlickewei, S-unit equations over number fields. *Invent. Math.* **102** (1990), 95–107.

Thanks!