Quasi-Automatic sequences

Jason Bell (joint with Boris Adamczewski)

We begin with Christol's theorem. This theorem gives a characterization of algebraic power series over finite fields. We say that $f(t) \in F_q[[t]]$ is *algebraic* if there exist polynomials $p_n, \ldots, p_0 \in F_q[t]$ with $p_n \neq 0$ such that

$$p_n(t)f(t)^n + \cdots + p_0(t) = 0.$$

To give Christol's characterization we must introduce the notion of a k-automatic sequence.

A sequence is k-*automatic* if its n'th term is generated by a finite state machine with n in base k as the input.

Examples of automatic sequences

The Thue-Morse sequence

$$0110100110010110100 10 \cdots$$

This sequence is 2-automatic.

e.g., n=13, then $n = 1101$ in base 2. Output $= 1$.

The Rudin-Shapiro sequence.

$$1, 1, 1, -1, 1, 1, -1, 1, 1, 1, 1, -1, -1, \ldots.$$

The n'th term of this sequence is given by -1 to the number of occurrences of "11" in the binary exmpansion of n.

This sequence is 2-automatic.

The *kernel* of a k-automatic sequence,

$$\{f(n) \mid n \geq 0\},$$

(or k-kernel) is the set of all subsequences of the form

$$\{f(k^a n + b) \mid n \geq 0\}$$

with $a \geq 0$ and $0 \leq b < k^a$.

Example: Take the Thue-Morse sequence

$$0110100110010110100 10\cdots.$$

If $TM(n)$ denotes the n'th term of the sequence, then

$$TM(2n) = TM(n) \qquad TM(2n{+}1) = 1{-}TM(n).$$

Thus either

$$TM(2^a n + b) = TM(n)$$

for all n, or

$$TM(2^a n + b) = 1 - TM(n)$$

for all n.

The 2-kernel of the Thue-Morse sequence consists of only two sequences; namely, the Thue-Morse sequence and its "opposite."

In general, we expect the k-kernel of a sequence to be infinite.

In the case of k-automatic sequences, however, the k-kernel is finite.

THEOREM: A sequence is k-automatic if and only if its k-kernel is finite.

An example from history?

- In the first century, Josephus along with 40 other rebels were hiding in a cave from the Romans during the Roman-Jewish war.
- Faced with certain death, the 41 men decided killing themselves was preferable to being killed by the Romans.
- Suicide was considered much worse than murder in Judaism (maybe–I don't really know anything about Judaism, to be honest).
- The men decided to form a circle and kill every other person in the circle till only one was left, the last person would then commit suicide.

Let $J(n)$ denote the last person to die in the Josephus circle of size n.

Then one sees that

$$J(2n) \ = \ 2J(n) - 1$$

$$J(2n+1) \ = \ 2J(n) + 1.$$

These relations show that for any k the sequence $\{J(n) \bmod k\}$ is 2-automatic.

Incidentally, when $n = 41$,

$J(41) = 2J(20) + 1 = 2(2J(10) - 1) + 1 = 19.$

Christol's Theorem says the following:

Let $f(t) = \sum_{i=0}^{\infty} a_i t^i \in F_q[[t]]$. Then $f(t)$ is algebraic if and only if the sequence (a_i) is p-automatic.

Idea behind proof: Let's look at $q = p$. Notice $a_i^p = a_i$ and

$$f(t)^{p^j} = \sum_{i=0}^{\infty} a_i t^{ip}.$$

The finiteness of the p-kernel allows us to deduce that f is algebraic.

There is a difference between the algebraic closure of $F((t))$ in the cases when F has positive characteristic and characteristic 0.

In the case that F has characteristic 0, we can say

$$\overline{F((t))} = \bigcup_n \overline{F}((t^{1/n})).$$

In characteristic $p > 0$, there is the following example of Artin and Schreier:

$$F(t)^p - F(t) - 1/t = 0.$$

Notice

$$F(t) = \sum_{i \geq 1} t^{-1/p^i}$$

is a root of this polynomial.

Because of this fact, Kedlaya looked at getting an analogue of Christol's theorem inside an algebraically closed field. To do this he uses the ring of *Hahn power series*

Give a field F, we let

$$F(t^{\mathbb{Q}})$$

denote the ring of power series

$$\sum_{\alpha\in\mathbb{Q}} c_\alpha t^\alpha$$

such that $S = \{\alpha \mid c_\alpha \neq 0\}$ is well-ordered; that is, there is no infinite decreasing subsequence $\alpha_1 > \alpha_2 > \cdots$ with $\alpha_i \in S$.

Why do we need the requirement that the support be well-ordered?

We define the product

$$\sum_{\alpha \in \mathbb{Q}} a_\alpha t^\alpha \sum_{\beta \in \mathbb{Q}} b_\beta t^\beta = \sum_{\gamma \in \mathbb{Q}} c_\gamma t^\gamma,$$

where

$$c_\gamma = \sum_{\alpha + \beta = \gamma} a_\alpha b_\beta.$$

It is necessary that these convolutions reduce to finite sums. We note that if the supports are well-ordered then the convolutions do indeed reduce to finite sums and, moreover, the set

$$\{\gamma \ : \ c_\gamma \neq 0\}$$

is well-ordered.

The Hahn power series (over an algebraically closed field) have the advantage of being algebraically closed and hence give a more natural setting for looking at algebraic power series. Kedlaya thus asked the question of whether a description similar to Christol's theorem for algebraic Hahn power series over finite fields could be given.

We now describe the work of Kedlaya and his generalization of ChristolÕs theorem. Let $k > 1$ be a positive integer. We set

$$\Sigma_k = \{0, 1, ..., k - 1, .\}$$

and we denote by

$$L(k)$$

the language on the alphabet Σ_k consisting of all words on Σ_k with exactly one occurrence of the letter Ô.Õ (the radix point) and whose first and last letters are not equal to 0.

We let S_k denote the set of nonnegative k-adic rationals; i.e.,

$$S_k = \{a/k^b : a, b \in Z, a \geq 0\}.$$

We note that there is a bijection between $L(k)$ to S_k.

To give his generalization of Christol's theorem, Kedlaya works with *quasi-automatic* maps. He defines

$$h : S_k \to \Delta$$

to be k-quasi-automatic if the support of h is well-ordered and there is a finite state machine which inputs a word W on $L(k)$ and outputs $h(w)$, where w is the element of S_k corresponding to W under the canonical bijection.

Kedlaya then proves the following theorem.

THEOREM: Let p be a prime, let q be a power of p, and let $f : Q \to F_q$. Then

$$\sum_{\alpha \in \mathbb{Q}} f(\alpha) t^{\alpha}$$

is algebraic over $F_q(t)$ if and only if there exist integers a and b with $a > 0$ such that $f(ax + b)$ is a p-quasi-automatic function.

Cobham's theorem and Christol's theorem.

Cobham's theorem states that a sequence that is both k- and ℓ-automatic is eventually periodic if k and ℓ are multiplicatively independent. This theorem has important consequences when combined with Christol's theorem.

Mahler asked the following question: Let (a_n) be a binary sequence and suppose that

$$\sum a_n/2^n$$

and

$$\sum a_n/3^n.$$

Is (a_n) necessarily eventually periodic? That is, are the numbers then both rational?

The corresponding question for power series is the following: If (a_n) is a binary sequence and

$$\sum [a_n]_2 t^n \in F_2[[t]]$$

and

$$\sum [a_n]_3 t^n \in F_3[[t]]$$

are both algebraic, are they necessarily both rational?

Note that Christol's theorem and Cobham's theorem show that this is the case.

One can ask whether there is a similar result for Hahn power series. The answer is 'yes' (almost) and it is easy.

The more difficult question is whether a k- and ℓ-quasi-automatic function is close to being eventually periodic. In the case that $S_k \cap S_\ell = \mathbb{Z}$, this result reduces to Cobham's theorem. On the other hand, it is possible for k and ℓ to be multiplicatively independent and yet have $S_k = S_\ell$ (e.g., $k = 6$, $\ell = 12$).

S-unit equations.

Let G be a finitely generated multiplicative subgroup of $\mathbb{Q}^*$. Given

$$(c_1, \ldots, c_d) \in \mathbb{Q}$$

one can ask how many solutions there are to the equation

$$\sum c_i g_i \;=\; 1,$$

$g_i \in G$ in which no proper subsum vanishes.

THEOREM: There are only finitely many such solutions.

This is the main result we need in obtaining the analogue of Cobham's theorem. We will be interested in particular in the case that G is generated by k and ℓ.

To do this, we first show that the support of a k-quasi-automatic subset of $S_k \cap [0, 1]$ is contained in a finite union of *Saguaro sets*. These are sets

$$S(W_1, \ldots, W_d; V_0, V_1, \ldots, V_d)$$

defined by all real numbers with base k expansion of the form

$$0.V_0 W_1^{i_1} V_1 W_2^{i_2} \cdots W_d^{i_d} V_d$$

The W's and V's are words on $\{0, 1, \ldots, k-1\}$.

It follows that for each Saguaro set there exist $c_1, \ldots, c_d$ such that every number in the set is of the form

$$\sum c_i k^{a_i}.$$

Consequently, to be both k- and ℓ-quasi-automatic, the elements of the support in $[0, 1]$ must be expressible in the form

$$\sum c_i k^{a_i}$$

and

$$\sum b_j \ell^{b_j}.$$

Using the multiplicative independence and the theorem on S-unit equations, we see that the support in $[0, 1]$ is finite.

We now glue the intervals together and deduce that if f is both k- and ℓ-quasi-automatic, then there exist a and b such that $f(ax+b)$ has support contained in $\mathbb{N}$ and is eventually periodic.

One can ask the additional questions such as: What happens if we remove the restriction that the support be well-ordered? In this case, the S-unit equation result does not apply since we cannot use the Saguaro set description. Nevertheless, the result appears to be true in this case.