8. The Skolem–Mahler–Lech Theorem

Jason P. Bell

Statement and history

One of Mahler's most important works was his contribution to the understanding of zeros of linear recurrences in characteristic zero; the culmination of this result is today known by the Skolem–Mahler–Lech theorem, and is a cornerstone in the theory of linear recurrence sequences. We recall that if K is a field then a K-valued *linear recurrence sequence* is a sequence $f : \mathbb{N}_0 \to K$ with the property that there exist $m \in \mathbb{N}$ and $b_1, \ldots, b_m \in K$ such that

$$f(n) = \sum_{j=1}^{m} b_j f(n-j)$$

for sufficient large natural numbers n. The Skolem–Mahler–Lech theorem asserts that if K has characteristic zero then a K-valued linear recurrence sequence $\{a_n\}_{n\in\mathbb{N}_0}$ has the property that the set of natural numbers n for which $a_n = 0$ is a union of at most finitely many infinite arithmetic progressions along with a finite set.

The first step in this theorem was made by Skolem [7], who proved the result when $a_n \in \mathbb{Z}$ for each n; Mahler's contribution [M31] was to extend Skolem's result to the case where a_n is an algebraic number for every n. Lech [4] extended the result to arbitrary fields of characteristic zero. (We note that Mahler [M138, M145] later gave a different, independent, proof from that of Lech of the general result. It is said that Mahler was unaware of Lech's result at the time of publishing his proof and was somewhat embarrassed by the discovery that Lech had proved the result already.)

Skolem's method and proof

One of the reasons for the huge impact of the Skolem–Mahler–Lech theorem is that linear recurrence sequences appear in many different guises, and this has led to many subsequent generalizations and extensions of the theorem. If K is a field (no assumption on characteristic) and $f : \mathbb{N}_0 \to K$ is a K-valued sequence then there are many equivalent formulations of being a linear recurrence. In fact, the following are equivalent:

1. the sequence $f(n)$ satisfies a linear recurrence;

2. there exist a natural number m, a matrix $A \in M_m(K)$, and column vectors $v, w \in K^m$ such that for $n \geq 0$ we have

$$f(n) = w^t A^n v;$$

3. the formal power series

$$\sum_{n \geq 0} f(n) x^n$$

is the power series expansion of a rational function (not having a pole at $x = 0$) about $x = 0$;

4. there exist $\ell \in \mathbb{N}$, $d \in \mathbb{N}_0$ and $c_{i,j}, \alpha_j \in \overline{K}$ for $0 \leq i \leq d$ and $1 \leq j \leq \ell$ such that for all sufficiently large n we have the formula

$$f(n) = \sum_{i=0}^{d} \sum_{j=1}^{\ell} c_{i,j} n^i \alpha_j^n.$$

The final characterization of linear recurrence sequences on the above list is what is exploited in the proof of the Skolem–Mahler–Lech theorem. The proof goes via what is today known as Skolem's method. Lech's key insight that allowed him to go beyond what Skolem and Mahler had accomplished was to observe that for a K-valued linear recurrence sequence with K of characteristic zero, one can restrict to a finitely generated extension of the rational numbers; namely, the field $\mathbb{Q}((c_{i,j})_{i,j}, \alpha_j))$, where the $c_{i,j}$ and α_j are as in item (4). It can then be shown, using the standard Lefschetz principle arguments, that if p is a prime number then $\mathbb{Q}((c_{i,j})_{i,j}, (\alpha_j)_j)$ can be embedded into a finite extension E of $\mathbb{Q}_p$; moreover this can be done in such a way that the nonzero elements among $\{c_{i,j}, \alpha_j\}$ are sent to units in the valuation ring of E. We remark that this embedding is immediate in the case that $f(n)$ takes values in algebraic numbers, since the field $\mathbb{Q}((c_{i,j})_{i,j}, (\alpha_j)_j)$ is a finite extension of $\mathbb{Q}$. Thus this embedding is the missing insight that Lech needed to extend Skolem's result and Mahler's result. Once one has this, it can then be shown that there exists a natural number k such that for each $b \in \{0, 1, \ldots, k-1\}$, the map

$$n \mapsto f(nk + b) = \sum_{i=1}^{d} \sum_{j=1}^{\ell} c_{i,j} (kn + b)^i \alpha_j^b (\alpha_j^k)^n$$

can be interpolated by a p-adic analytic map. Since the valuation ring of E is compact, one can then use Strassman's theorem and conclude that for a given b, the map either vanishes identically on E or it has only finitely many zeros. Translating this statement to the linear recurrence, we see that either $f(kn + b)$ is either identically zero for all n or there are finitely many zeros on

the progression $k\mathbb{N} + b$. We remark that since characterization (4) generally only applies to large n, it is possible that $f(kn + b) = 0$ for sufficiently large n rather than all n, but this still gives the same conclusion about arithmetic progressions along with finite sets.

Proposed extensions of the Skolem–Mahler–Lech theorem

In this section, we describe two proposed extensions of the Skolem–Mahler–Lech theorem: one dynamical, and one involving differential equations.

We begin with a conjecture of Rubel. We recall that a characterization of a linear recurrence sequence over a field K is that it is the power series expansion of a rational function. Rational functions form a distinguished subset of the collection of so-called D-finite power series; these are power series which satisfy a homogeneous linear differential equation with polynomial coefficients. The class of D-finite power series is very broad, including power series expansions of algebraic functions along with exponentials and other important functions. Rubel [6, Problem 16] conjectured that a complex power series

$$f(z) := \sum_{n=0}^{\infty} a_n z^n,$$

satisfying a homogeneous linear differential equation with polynomial coefficients, should have the property that the set

$$S := \{n \in \mathbb{N}_0 \colon a_n = 0\}$$

has the same form as in the Skolem–Mahler–Lech theorem: namely that it should be a finite union of infinite arithmetic progressions along with a finite set. This result would have substantial implications for Diophantine equations, since many such equations can be expressed in terms of zero sets of coefficients of D-finite power series.

In terms of progress on Rubel's conjecture, the strongest current result is due to Bézivin [2] and Methfessel [5], who showed using Szemerédi's theorem that a weak version holds: the zero set is a finite union of infinite arithmetic progressions along with a set of zero density.

We now give a proposed dynamical extension of the Skolem–Mahler–Lech theorem. We note that if one uses the matrix characterization of linear recurrence sequences given above, then one can view a linear recurrence sequence $f(n)$ as a sequence of the form $w^T A^n v$ where A is a matrix with entries in K and v and w are column vectors. Then $f(n) = 0$ if and only if $A^n v$ lies in the subspace $w^\perp$ of vectors x such that $w^T \cdot x = 0$. If we think of A as a linear endomorphism of some vector space V then $\{A^n v \colon n \geq 0\}$ is just the orbit of the vector v under this linear transformation. Thus the Skolem–Mahler–Lech theorem asserts that the set of points in this orbit that lie in a given subspace

can be described in terms of arithmetic progressions along with a finite set. We note that if A is invertible then one can "run the recurrence backwards" and Skolem's method then gives that the set of integers for which $w^T A^n v = 0$ is a finite union of doubly-infinite arithmetic progressions along with a finite set. One way to think of this is that the zero set is then a finite union of cosets of additive subgroups of $\mathbb{Z}$ (with a singleton being a coset of the trivial group). This immediately reminds one of the Mordell–Lang conjecture (now a theorem due to Faltings and Vojta), which states that in a semiabelian variety X (we use addition as the group operation) over $\mathbb{C}$, if Γ is a finitely generated (abelian) subgroup and Y is a Zariski closed subset of Y then $\Gamma \cap Y$ is a finite union of cosets of subgroups of Γ. This result can also be recast in a dynamical fashion. For each $\gamma \in \Gamma$, we have a translation map T_γ on X, $x \mapsto \gamma + x$. If Γ is an infinite cyclic group with generator γ, then the result is simply that $\{n \colon T_\gamma^n(e) \in Y\}$ is a finite union of cosets of Γ.

As it turns out, the Skolem–Mahler–Lech theorem and the cyclic case of the Mordell–Lang conjecture can be unified into a single dynamical conjecture, which is called the Dynamical Mordell–Lang conjecture. This conjecture states that given an endomorphism Φ of a quasiprojective variety X defined over an algebraically closed field K of characteristic zero, a point $\alpha \in X(K)$, and a subvariety V of X, the set of n for which $\Phi^n(x) \in V(K)$ is a finite union of arithmetic progressions along with a finite set. When Φ is an étale self-map the result is known [1]. In particular, this gives the case when Φ is an automorphism and one can derive both the cyclic case of the Mordell–Lang conjecture and the Skolem–Mahler–Lech result form this theorem.

1 The Skolem–Mahler–Lech Theorem in positive characteristic

The Skolem–Mahler–Lech theorem naturally leads to the question of what happens for zero sets of linear recurrence sequences in positive characteristic? In this setting, Lech gave examples which show that the direct translation of the theorem does not hold in characteristic $p > 0$. The simplest counterexample is to take $K = \mathbb{F}_p(t)$ and take $f(n) = (1+t)^n - t^n - 1$, which is easily seen to be a linear recurrence sequence. Then if n is a power of p, we see that $f(n) = 0$; on the other hand, if n is not a power of p then there is some $j \in \{1, 2, \ldots, n-1\}$ such that $\binom{n}{j}$ is nonzero mod p and since t is transcendental we then see that $f(n)$ is nonzero. Thus the zero set of $f(n)$ is just the sequence $\{1, p, p^2, \ldots\}$, which is not a finite union of arithmetic progressions along with a finite set.

A positive characteristic analogue of the Skolem–Mahler–Lech theorem was given by Derksen [3]. Derksen's analogue of the Skolem–Mahler–Lech theorem shows that the zero sets of linear recurrences over a field of characteristic $p > 0$ are given by finite unions of arithmetic progressions along with finite sets along with what he names *p-normal* sets. Intuitively, p-normal sets are sets built up from powers of p, much like the set in the counterexample just given.

Although the positive characteristic characterization is somewhat more complicated than the elegant characterization given in the Skolem–Mahler–Lech theorem (characteristic zero), Derksen's result has one key advantage: it is effective. Skolem's problem is whether it is decidable if a $\mathbb{Z}$-valued linear recurrence sequence has a zero term. This problem is still open, despite its simple-looking nature. In positive characteristic, Derksen is able to provide algorithms which allow one to completely settle the positive characteristic analogue of Skolem's problem and, moreover, to describe the zero set completely. Much of Derksen's argument relies on using *finite-state automata*, from which he is able to derive effectivity.

We recall that if p is a prime number and Δ is a finite set, then one has the notion of a p-automatic map

$$f : \mathbb{N}_0 \to \Delta,$$

which is defined as follows. For each $j \in \{0, 1, \ldots, p-1\}$, we have a map

$$e_j : \mathbb{N}_0 \to \mathbb{N}_0$$

defined by $e_j(n) = pn + j$. We let Σ denote the submonoid of self-maps on $\mathbb{N}_0$ generated by the collection of all e_j under composition. The map f is said to be a *p-automatic sequence* if there are only finitely many distinct sequences in the collection

$$\{f \circ e \ : \ e \in \Sigma\}.$$

Then we can say that a subset $S \subseteq \mathbb{N}_0$ is a *p-automatic set* if the characteristic sequence of S is a p-automatic sequence.

The name automatic comes from the fact that such sequences can in fact be generated by a finite state automaton, taking as input the base-p expansion of a natural number n as input and giving $f(n) \in \Delta$ as output. Derksen's first result towards his characterization was that if $f(n)$ is a sequence satisfying a recurrence over a field K of characteristic p, then the set of natural numbers n such that $f(n) = 0$ is a p-automatic set. Even this first theorem is enough to deal with the thorny decidability problems that one encounters in characteristic zero, as his proof gives a way of bounding the number of states in the corresponding automaton.

In fact, Derksen gives a refinement of this result by describing which types of automata can occur; by doing so, he obtains his characterization of p-normal sets. Let p be a prime number. A subset $S \subseteq \mathbb{N}$ is *p-normal* if it is a finite union of sets of the form

$$S_q(c_0; \ldots, c_d) := \{c_0 + c_1 q^{k_1} + \cdots + c_d q^{k_d} : k_1, \ldots, k_d \in \mathbb{N}\} \cap \mathbb{N},$$

where q is a power of p and $c_0, \ldots, c_d$ are nonzero rational numbers satisfying $c_0 + \cdots + c_d \in \mathbb{Z}$ and $(q-1)c_i \in \mathbb{Z}$, for $i = 0, \ldots, d$.

After obtaining his characterization of zero sets in terms of automata, Derksen shows that if one has a linear recurrence sequence that does not vanish on any infinite arithmetic progressions then the automata which accepts the natural numbers whose base-p expansions are in the zero set has the property that it cannot have two distinct cycles based at a single state. This in turn gives Derksen's main theorem: if $f(n)$ is a sequence satisfying a recurrence over a field K of characteristic p, then the set of natural numbers n such that $f(n) = 0$ is a finite union of arithmetic progressions along with a p-normal set.

REFERENCES

[M31] K. Mahler, *Eine arithmetische Eigenschaft der Taylor-Koeffizienten rationaler Funktionen*, Proc. Akad. Wet. Amsterdam 38 (1935), 50–60.

[M138] K. Mahler, *An interpolation series for continuous functions of a p-adic variable*, J. Reine Angew. Math. (Crelle) 199 (1958), 23–34.

[M145] K. Mahler, *A correction to the paper "An interpolation series for continuous functions of a p-adic variable"*, J. Reine Angew. Math. (Crelle) 208 (1961), 70–72.

[1] J. P. Bell, D. Ghioca, and T. J. Tucker, *The dynamical Mordell–Lang problem for étale maps*, Amer. J. Math. 132 (2010), no. 6, 1655–1675.

[2] J.-P. Bézivin, *Une généralisation du théorème de Skolem–Mahler–Lech*, Quart. J. Math. Oxford Ser. (2) 40 (1989), no. 158, 133–138.

[3] H. Derksen, *A Skolem–Mahler–Lech theorem in positive characteristic and finite automata*, Invent. Math. 168 (2007), no. 1, 175–224.

[4] C. Lech, *A note on recurring series*, Ark. Mat. 2 (1953), 417–421.

[5] C. Methfessel, *On the zeros of recurrence sequences with non-constant coefficients*, Arch. Math. (Basel) 74 (2000), no. 3, 201–206.

[6] L. A. Rubel, *Some research problems about algebraic differential equations*, Trans. Amer. Math. Soc. 280 (1983), no. 1, 43–52.

[7] T. Skolem, *Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen und diophantischer Gleichungen*, Comptes Rendus Congr. Math. Scand. (Stockholm, 1934) (1934), 163–188.

Jason P. Bell
Dept. of Pure Mathematics
University of Waterloo
Waterloo, ON N2L 3G1
Canada