The Berstel-Mignotte Result Revisited

Bruce Litow: School of MPIT, James Cook University
Townsville, Queensland Australia

In 1930, Alfred Tarski devised a decision method for the first order theory of real fields by quantifier elimination, although he did not write up a formal exposition until 1948.

This theory, interpreted over the reals is called Tarski algebra (TA), and the current best upper bound on quantifier elimination time is

Theorem 1.1 *A prenex TA formula of size n, in m variables and a quantifier alternations can be converted into an equivalent quantifier-free formula in time*

$$n^{m^{O(a)}}.$$

In 1976, Berstel and Mignotte proved that it is decidable to determine whether infinitely many terms of a linear recurrence over $\mathbb{Q}$ are 0. Actually, they rely on a result external to their paper, which is the basis for decidability. Berstel and Mignotte are mainly concerned with deriving an upper bound on the LCM of the degrees of periodic zeros associated to the recurrence.

We want to make explicit some of the time bounds, and also the core of the decision method.

What is it?

Theorem 1.2 (Skolem-Mahler-Lech) *The set of indices i for which $a_i = 0$ is called the zero set of the sequence a_i. Over any field of characteristic zero, the zero set of a linear recurrence is ultimately periodic (U.P., finite union of arithmetic progressions).*

The assertion can fail over positive characteristic fields. (Lech).

Skolem-Pisot Problem. Over $\mathbb{Q}$ is their an algorithm to decide whether the zero set of a linear recurrence is empty? Answer: unknown.

We look at Skolem-Pisot in terms of rational series P/Q, where $Q(0) = 1$. That is, is $[n](P/Q) = 0$ for some n ?

Classical fact: Letting $\alpha_1, \ldots, \alpha_s$ be the distinct zeros of Q, Skolem-Pisot reduces to asking whether forms like

$$\sum_{i=1}^{s} C_i \cdot \alpha_i^n = 0$$

for some n. Here, C_i is a polynomial in n with coefficients in $\mathbb{Q}[\alpha_i]$.

Some results. On the positive side: For recurrences of rank at most 5 their is an algorithm. See TR-TUCS (Turku) 683, April 2005 Halava-Harju-Hirvensalo-Karhumäki.

On the negative side: The analogue of Skolem-Pisot for multivariate rational series (P/Q where P, Q are multivariate) is known to be undecidable for 9 or more variables by reduction from Hilbert's 10th problem.

The reduction proceeds via a mapping from a k-variate Diophantine polynomial P to a k-variate rational series f_P with the property that for $c_1, \ldots, c_k \in \mathbb{N}$, $P(c_1, \ldots, c_k) = [c_1, \ldots, c_k]f_P$. It is clear that the solution set of $P = 0$ is the complement of the set of support of f_P.

Salomaa and Soittola have shown how to reduce Hilbert's 10th problem to formal rational series in two noncommuting variables. However, their construction does not work for commuting variables because it uses closure under Hadamard product for rational noncommuting multivariate series (essentially Schützenberger-Jungen), and this closure fails for the commuting variable case.

First compute f_P when P is the monomial $x_1^{g_1} \cdots x_k^{g_k}$. Let ∂_i denote partial differentiation by x_i followed by multiplication by x_i.

Define f_P to be

$$\partial_1^{g_1} \cdots \partial_k^{g_k} \frac{1}{(1-x_1)\cdots(1-x_k)} \ .$$

Check that

$$[c_1, \ldots, c_k] f_P = c_1^{g_1} \cdots c_k^{g_k} \ .$$

Note that ∂_i^0 is the identity operator. It is clear that f_P is rational.

If P is a Diophantine polynomial and $b \in \mathbb{Z}$, define $f_{b \cdot P} = b \cdot f_P$. If $P = R + S$, where R, S are Diophantine polynomial, define $f_P = f_R + f_S$. These definitions satisfy the required property. Since any Diophantine polynomial can be built up by a finite number of applications of addition of polynomials and multiplication by an integer, starting from monomials, we see that f_P is always a rational series.

A Warm-up Exercise

We know that there are at least four ways to deal with problems of the form

$$A = \sum_{i=1}^{r} \alpha_i = 0 \ ?$$

where the α_i are algebraic numbers.

- ** Tarski Algebra (TA).
- Elimination methods.
- Gröbner basis methods.
- Direct constructions.

We assume that each α_i has degree at most d.

One way to decide whether $A = 0$ is to construct an algebraic number β and rational coefficient polynomials P_i such that $\alpha_i = P_i(\beta)$.

Then, $A = 0 \Leftrightarrow P(\beta) = 0$ where P is a rational coefficient polynomial. Now, this equation holds $\Leftrightarrow P = R \cdot S$, where R is the minimal polynomial (which we know because that is how we have specified β). Whether $P \equiv 0 \bmod R$ is certainly computable.

Specialize $A = 0$ to roots of unity, $\alpha_i = \omega_i^{a_i}$, where $\mathbf{e}(z) = \exp(2\pi\iota \cdot z)$, and ι is the positive branch of $\sqrt{-1}$. Let $a_i < b_i$ be non negative integers. We know that b_i is bounded above by $d^{O(1)}$.

How hard is it to determine whether $A = 0$?

Let's look at two approaches.

Approach 1.

Letting $B = \mathrm{LCM}(b_1, \ldots, b_r)$, and $B_i = B/b_i$,

$$A = 0 \Leftrightarrow \sum_{i=1}^{r} \omega^{b_i} = 0 \ ,$$

where $\omega = \mathbf{e}(1/B)$. Specializing our method to roots of unity,

$$A = 0 \Leftrightarrow P \equiv 0 \bmod R \ ,$$

where $P = \sum_{i=1}^{r} x^{B_i}$, and R is the cyclotomic polynomial fr ω (degree $\phi(B)$).

Brute force shows that R can be computed in $O(B^2)$ time (one can do better), and from this $A = 0$ is also in $O(B^2)$ time for our roots of unity.

Approach 2.

Determining whether $A = 0$ is in $d^{O(1)}$ time. Approximate each summand by its Taylor series truncated to enough terms so that the error is bounded above by $1/(2r \cdot d^d)$. Clearly, the number of terms is bounded by $d^{O(1)}$. Let A_* be the sum of these truncated series, so that $|A - A_*| < 1/(2 \cdot d^d)$. If $|A_*| \geq 1/(2 \cdot d^d)$, then $A \neq 0$. If $|A_*| < 1/(2 \cdot d^d)$, then $|A| < 1/d^d$, which implies $A = 0$. These tests can be carried out in $d^{O(1)}$ time.

Let $A(n) = \sum_{i=1}^{r} C_i \cdot \alpha_i^n$. Asking whether $A(n) = 0$ for some n is the core of SP. If the α_i are roots of unity as before, then the $d^{O(1)}$ method extends to the case with the coefficients C_i.

Berstel and Mignotte showed that $B = \exp(O(\sqrt{d \log d}))$, so that SP for roots of unity can be solved in time

$$\exp(O(\sqrt{d \log d})) \cdot d^{O(1)} = \exp(O(\sqrt{d \log d})) \ .$$

In fact, this bound extends to SP generally.

We classify the zeros of Q as periodic or non periodic. Recall that a complex number α is periodic if $\alpha/|\alpha|$ is a root of unity, otherwise it is non periodic.

It is easy to decide for a given algebraic number α whether it is periodic. In fact, α is periodic $\Leftrightarrow (\alpha/|\alpha|)^j = 1$ for some $\phi(j) \leq 2d^2$, where d is the degree of α.

By Theorem 1.1 periodicity is in $\mathbf{P}$ ($(O(1)$ variables).

- If β has degree k, so does $1/\bar{\beta}$.
- If β and γ have degrees j, k, respectively, then $\beta\gamma$ has degree at most $j \cdot k$.
- If β has degree d, then each branch of $\sqrt{\beta}$ has degree at most $2d$.
- $(\alpha/|\alpha|)^2 = \alpha/\bar{\alpha}$, so $\alpha/|\alpha|$ has degree at most $2d^2$. The bound on j follows from the fact that if ω is a root of unity of degree k, then $\omega^j = 1$ for some $\phi(j) \leq k$.

We now show that SP_∞ $d^{O(1)}$-time reduces to the roots of unity case. This is done by:

1. If a recurrence P/Q has infinite zero set, then for each aperiodic zero α_i of Q, $C_i = 0$.
2. Whether α_i is aperiodic is in $d^{O(1)}$ time.
3. Determining whether $C_i = 0$ is in $d^{O(1)}$ time.

We have already taken care of item 2. Item 3 is also follows from Theorem 1.1. It requires a careful description of the C_i. We look more closely at Item 1.

Recall that complex numbers $\beta_1, \ldots, \beta_s$ are said to be linearly independent over $\mathbb{Q}$ (LIQ) if for rational $a_0, \ldots, a_s$,

$$a_0 + a_1 \cdot \beta_1 + \cdots + a_s \cdot \beta_s = 0 \Leftrightarrow a_0 = \cdots = a_s = 0 \; .$$

The Weyl-von Neumann Theorem:
If $\beta_1, \ldots, \beta_s$ are LIQ, then

$$\{(n \cdot \beta_1 \bmod, \ldots, n \cdot \beta_s \bmod 1) \mid n \in \mathbb{N}\}$$

is dense in the interval $[0, 1]$.

This implies that

$$\{(\mathbf{e}(n \cdot \beta_1), \ldots, \mathbf{e}(n \cdot \beta_s)) \mid n \in A\} \ ,$$

where A is a U.P. set, is dense on the unit circle.

We can now see why aperiodic zeros are excluded from playing any role in the zero set of a linear receurrence. Assume a recurrence has infinite zero set K, and there are two sets of aperiodic zeros, $\beta_1, \ldots, \beta_s$ and $\gamma_1, \ldots, \gamma_t$, such that $\beta = |\beta_i|$, $\gamma = |\gamma_j|$, and $\gamma < \beta$.

Write $\beta_i = \beta \cdot \mathbf{e}(\mu_i)$, and $\gamma_j = \gamma \cdot \mathbf{e}(\nu_j)$.
Assume further that $\mu_1, \ldots, \mu_s$ are LIQ, and $\nu_1, \ldots, \nu_t$ are LIQ.

For $n \in K$, the n-th recurrence term looks like:

$$(\beta/\gamma)^n \cdot (\sum_{i=1}^{s} a_i \cdot \mathbf{e}(n \cdot \mu_i)) + \sum_{j=1}^{s} b_j \cdot \mathbf{e}(n \cdot \mu_j) \ .$$

By Weyl-von Neumann, there is some $\delta > 0$ such that for n sufficiently large,

$$|\sum_{i=1}^{s} a_i \cdot \mathbf{e}(n \cdot \mu_i)| > \delta \ .$$

Letting

$$D = \sum_{j=1}^{t} |b_j| \ ,$$

we can pick $n \in K$ so large that

$$(\beta/\gamma)^n \cdot \delta > D \ ,$$

which implies that this term can not vanish.

A couple of technical points.

1. Coefficients are polynomials in n. However, this requires no new ideas.
2. Among aperiodic zeros of given modulus there may be some whose arguments e.g. some μ_i, may be linearly dependent on others. This requires a more careful argument. Details are in the paper. be

Complexity of SP_∞ reduces to the 'roots-of-unity' variant of SP.
This appears to be an open problem.