ZEROES OF INTEGER LINEAR RECURRENCES

DANIEL LITT

1. Introduction

Consider the integer linear recurrence

$$x_n = x_{n-1} + 2x_{n-2} + 3x_{n-3}$$

with

$$x_0 = x_1 = x_2 = 1.$$

For which n is $x_n = 0$?

Answer: x_n is never zero, because it's always positive.

What about

$$x_n = -2x_{n-1} + x_{n-2}$$

with $x_0 = 2, x_1 = 1$?

Answer: Only $n = 3$, because

$$x_n = \left(1 - \frac{3}{2\sqrt{2}}\right)(-1-\sqrt{2})^n + \frac{1}{4}(4+3\sqrt{2})(\sqrt{2}-1)^n$$

and the left term dominates for large n. (Discuss when this will work, drawing roots of polynomial.)

What about

$$x_n = 2x_{n-1} - 3x_{n-2}$$

with $x_0 = 0, x_1 = 1$?

Answer: Only x_0; look mod 3. (Absolute value doesn't work, because roots are $1 \pm i\sqrt{2}$ which have the same absolute value.)

Mod 3 this is $(-1)^{n+1}$.

Finally, how about

$$x_n = x_{n-2},$$

with $x_0 = 0, x_1 = 1$?

Answer: Even n, obviously.

The goal of this talk is to show that these sorts of behaviors are the only thing that can happen really; we'll prove a beautiful theorem of Skolem-Mahler-Lech to that effect.

Theorem 1 (Skolem-Mahler-Lech)**.** *Let a_n be a sequence defined by an integer linear recurrence. Then the set*

$$A := \{n \mid a_n = 0\}$$

is the union of a finite set and a finite collection of (right)-infinite arithemtic progressions, i.e. sets of the form

$$\{qn + r \mid n \in \mathbb{N}\}.$$

2. The p-adics

Let's look at the example

$$x_n = 2x_{n-1} - 3x_{n-2}$$

more closely. We have

$$x_n = -\frac{1}{2\sqrt{-2}}(1-\sqrt{-2})^n + \frac{1}{2\sqrt{-2}}(1+\sqrt{-2})^n.$$

I've purposefully written this so it makes sense mod 3. Now $\sqrt{-2} = \pm 1 \bmod 3$, giving the mod 3 identity I claimed above.

This is rigorous because -2 has a square root mod 3; of course looking mod 3 might not have worked; we might have had to look mod a higher power; for example, if our initial conditions were $x_0 = 0, x_1 = 3$. But that's ok–we can lift our roots mod 9. Namely, pick an arbitrary lift of one of our roots, say 1; now we want to write $(3a+1)^2 = -2 \bmod 9$, or $6a + 1 = -2 \bmod 9$, which has a solution (namely $a = 1$). Why does this work?

Well in general, lets say we have a root x of some polynomial p mod 3. We can pick some lift $\bar{x}$ mod 9 and try to fix it up, e.g. we want $p(3a + \bar{x}) = 0 \bmod 9$ for $a \in \{0, 1, 2\}$. In other words

$$3p'(\bar{x})a + p(\bar{x}) = 0 \bmod 9$$

But $p(\bar{x}) = 0 \bmod 3$ by our choice of $\bar{x}$, so a solution exists for a if $p'(\bar{x})$ is invertible; that is, if $p'(x) \neq 0$. By the way, this is just Newton's method! Of course, nothing is special about 3 and 9, so we've really just proved

Lemma 1 (Hensel's Lemma)**.** *Let f be a polynomial with integer coefficients and p a prime. Suppose x is a root of f mod p, and $f'(x) \neq 0$. Then the exists $\bar{x} \in \mathbb{Z}/p^k\mathbb{Z}$ so that $f(x) = 0$ and $\bar{x} = x \bmod p$.*

Proof. Induction on k. □

I claim this method of analyzing linear recurrences—looking mod large powers of primes—is secretly the same as the previous method using absolute values. Namely, let's consider this cleaner statement of Hensel's lemma. Define the p-adic integers

$$\mathbb{Z}_p := \varprojlim \mathbb{Z}/p^k\mathbb{Z}.$$

An element of $\mathbb{Z}_p$ is by definition a sequence of elements $a_k \in \mathbb{Z}/p^k\mathbb{Z}$ so that $a_k = a_{k-1} \bmod p^{k-1}$.

Then Hensel's Lemma says that if a polynomial f with integer coefficients (or really with coefficients in $\mathbb{Z}_p$!) has a $x \in \mathbb{F}_p$ with $f'(x) \neq 0$, then f has a solution $\bar{x}$ in $\mathbb{Z}_p$ so that $\bar{x} = x \bmod p$.

$\mathbb{Z}_p$ is a topological ring, with multiplication inherited from the finite levels and a topology induced by giving each $\mathbb{Z}/p^k\mathbb{Z}$ the discrete topology. (So open sets are generated by preimages of subsets of $\mathbb{Z}/p^k\mathbb{Z}$ for some k.) [Draw a picture of $\mathbb{Z}_3$; note that any point in a ball is its center]

Now each p-adic number has a "decimal" expansion

$$z = \sum_{k \geq 0} a_k p^k$$

where $p \in \{0, 1, 2, \cdots, p-1\}$ by the explicit description of the inverse limit; note that those numbers with finite decimal expansions are just the usual integers. Just like the usual decimal expansion of a number, this should suggest another description of the p-adics. Namely, if z is a non-zero integer, define

$$v_p(z) = \max\{k \in \mathbb{N} \mid p^k \text{ divides } z\}.$$

Let

$$|z|_p = p^{-v_p(z)}$$

for z nonzero, and let $|0|_p = 0$. So for example $|24|_2 = 2^{-3} = 1/8$. This is an absolute value in the sense that it satisfies:

- $|0|_p = 0$;
- $|x|_p|y|_p = |xy|_p$;
- $|x|_p + |y|_p \geq |x + y|_p$; indeed $\max(|x|_p, |y|_p) \geq |x + y|_p$. (This is called the ultrametric inequality).

This absolute value induces a metric on $\mathbb{Z}$; its completion with respect to the induced topology is $\mathbb{Z}_p$! The extension of v_p to $\mathbb{Z}_p$ is that

$$v_p\left(\sum_k a_k p^k\right) = \min(k \mid a_k \neq 0).$$

Alright, so let's go back to our example

$$x_n = 2x_{n-1} - 3x_{n-2}.$$

The formula

$$x_n = -\frac{1}{2\sqrt{-2}}(1-\sqrt{-2})^n + \frac{1}{2\sqrt{-2}}(1+\sqrt{-2})^n.$$

makes sense in $\mathbb{Z}_3$ by Hensel's lemma, where we take $\sqrt{-2} = -1 \bmod 3$. But now

$$|1-\sqrt{-2}|_3 = 1, \text{ and } |1+\sqrt{-2}|_3 = 1/3$$

so we may apply our old absolute value argument!

3. Topology and Analysis in the p-adics

Let's establish some properties of the p-adics. I've already asserted their completeness; one way to see this is that if b_k is a Cauchy sequence, the i-th digit of b_k stabilizes (why?), giving the i-th digit of the limit. Having convergence of sequences, let us move on to convergent series.

Lemma 2. *Let b_k be a sequence of p-adic integers. Then*

$$\sum b_k$$

converges iff $|b_k|_p \to 0$.

Proof. By the ultrametric inequality, the partial sums form a Cauchy sequence iff the $|b_k|_p \to 0$. □

We can now do analysis in the p-adics.

Definition 1. *Let B be an open ball in $\mathbb{Z}_p$. A function $f : B \to \mathbb{Z}_p$ is p-adic analytic if it is defined by a power series*

$$f(z) = \sum_{k\geq 0} a_k (z-b_0)^k$$

for some $b_0 \in B$, with the power-series convergent for all $z \in B$.

There are two natural ways to differentiate such a power series f. For simplicity, let's assume it's centered at zero and has radius of convergence at least 1 (namely it converges on all of $\mathbb{Z}_p$). Then we may formally differentiate f via

$$f(z) = \sum_k a_k z^k; f'(z) = \sum_k k a_k z^{k-1}.$$

Note that the radius of convergence of f' will be at least that of f, as differentiating has made the coefficients no larger in p-adic absolute value.

On the other hand, of $s_i \to 0, s_i \neq 0$, one may define f' via a difference quotient:

$$f'(z) = \lim_{i\to\infty} \frac{f(z+s_i)-f(z)}{s_i}.$$

It's not hard to check that these two definitions agree, say by looking at monomials and using continuity of evaluation. In particular, this implies that a p-adic analytic function is infinitely differentiable, and is constant if and only if its derivative is identically zero.

This lets us do things like take Taylor series, recenter our analytic function, and so on. In particular, an analytic function with radius of convergence R about b_0 will be analytic with radius of convergence R about any b_1 with $|b_0 - b_1|_p \leq R$.

We'll now prove a beautiful theorem of Strassmann, which will be the main technical tool in our proof of the Mahler-Skolem-Lech theorem. This result is analogous to the fact that a holomorphic function on a connected region is identically zero if it is zero on a set with limit points.

Theorem 2 (Strassmann)**.** *Let $f : B \to \mathbb{Z}_p$ be a p-adic analytic function. Then either f is identically zero, or has only finitely many zeros in B.*

Lemma 3. *$\mathbb{Z}_p$ is compact.*

Proof. Let $\{b_k\}$ be a sequence of elements of $\mathbb{Z}_p$; we wish to show there is a convergent subsequence. Let a_{ik} be the i-th "digit" of b_k. Then the a_{1k} take some fixed value for infinitely many indices k_j; among these k_j, the a_{2k_j} take some value infinitely often, and so on. Collecting all of these values gives the decimal expansion of a limit point. □

Proof of Strassmann's theorem. Suppose f has infinitely many zeros, say $f(b_k) = 0$. Then by our lemma above, the b_k have a limit point b. Let expand f about b via

$$f(z) = \sum a_k (z-b)^k.$$

If f is not identically zero, some $a_k \neq 0$; let a_N be the first such coefficient. Then

$$f(z) = (z-b)^N (a_N + (z-b)g(z)),$$

and so for $|z-b|_p$ very small, we have that

$$|(z-b)g(z)|_p < |a_N|_p.$$

In particular, f is non-zero in some small punctured disk about b. But this contradicts the fact that b was a limit point of the b_k. □

4. Interlude: There are infinitely many primes

Consider the following topology on the integers: A basis is given by sets of the form

$$S(a,b) := a\mathbb{Z} + b, a \neq 0.$$

Note that the $S(a,b)$ are both open and closed, and any non-empty open set is infinite. Using the fact that this is a topology, we may prove the following beautiful theorem:

Theorem 3. *There are infinitely many primes.*

Proof. Assume the contrary. Then there are finitely may primes p_i, and

$$\bigcup_i S(p_i, 0) = \mathbb{Z} \setminus \{\pm 1\}.$$

In particular, $\{\pm 1\}$ is open, which is a contradiction. □

(This is of course just Euclid's proof in disguise.) Why do I mention this? Well, the topology I've described on $\mathbb{Z}$ is exactly the subspace topology induced by the inclusion

$$\mathbb{Z} \to \prod_p \mathbb{Z}_p.$$

This should suggest that p-adic analysis might have something to say about arithmetic progressions. Namely, let's imagine for a second that given an integer linear recurrence sequence a_k we could find a collection of p-adic analytic functions f_i, $0 \leq i \leq m-1$ such that

$$f_i(n) = nm + i$$

for large n. Then each f_i would either be identically zero, or would have only finitely many zeroes, which would complete the proof. The goal is to find such functions; I think this is a natural goal, given that $\mathbb{Z}$ is dense in $\mathbb{Z}_p$, and we like to extend functions (in this case $k \mapsto a_k$) on dense subsets to their closures.

5. The Mahler-Skolem-Lech Theorem

Let me remind you what we're trying to prove.

Theorem 4 (Skolem-Mahler-Lech)**.** *Let a_n be a sequence defined by an integer linear recurrence. Then the set*

$$A := \{n \mid a_n = 0\}$$

is the union of a finite set and a finite collection of (right)-infinite arithemtic progressions, i.e. sets of the form

$$\{qn + r \mid n \in \mathbb{N}\}.$$

Proof. Let's rewrite our problem to try to find f_i as above. An integer linear recurrence sequence a_k may be written as

$$a_k = \langle A^k v, w\rangle$$

for A an $n \times n$ integer matrix, v, w vectors with integer entries (here v is the initial conditions; A is the transition matrix; and w picks out the top entry of a vector). Now choose some p not dividing the determinant of A; in particular A is invertible mod p.

Now the group of invertible matrices mod p is finite, so in particular $A^m = 1 \bmod p$ for some m; this will eventually be the period of our arithmetic progressions. Let us write $A^m = I + pB$ for some matrix B with integer entries. Now let us construct our f_i. For $0 \leq i \leq m-1$, we have

$$f_i(n) = a_{mn+i} = \langle A^{mn} A^i v, w\rangle = \langle (I + pB)^n A^i v, w\rangle.$$

This is looking good! Expanding $(1 + pB)^n$ via the binomial formula, we find that

$$f_i(n) = \sum_k p^k P_k(n)$$

for some polynomials P_k in n with integer coefficients (these come from binomial coefficients). This power series makes sense as a p-adic analytic function convergent on all of $\mathbb{Z}_p$!

But now we're done—each f_i has only finitely many zeros or is identically zero by Strassmann's theorem. □

This proof is due to Hansel.

Note that we've effectively bounded the period of the zero sets, namely by the order of $GL_n(\mathbb{F}_p)$, where n is the length of the recurrence relation and p is the smallest prime not dividing $\det(A)$ above. It is not hard to generalize this argument to a similar result over any domain whose fraction field is characteristic zero. The theorem fails dramatically in finite characteristic. For example, consider over the field $\mathbb{F}_p(t)$ the recurrence

$$x_n = (t+1)^n - t^n - 1$$

defined by

$$x_n = (2x+2)x_{n-1} - (x^2 + 3x + 1)x_{n-2} + (x^2 + x)x_{n-3}.$$

Then the zero set is exactly the p-th powers.

On the other hand, both the number and size of the exceptional zeros are mysterious from this proof. Evertse, Schlickewei, and Schmidt give a difficult bound on the number of exceptional zeros, (essentially by analyzing the f_i), but there is (as of 2007, and I assume now) no effective bound on their size. In particular, the following is open

Question 1. *Is there a computer program that decides whether or not an integer linear recurrence has any zeros?*