A Universal Skolem Set of Positive Lower Density

F. Luca, J. Ouaknine, J. Worrell

MFCS 2022, Vienna

Halting Problem for Single-Path Linear Loops

What is the simplest class of programs for which decidability of the Halting Problem is open?

Halting Problem for Single-Path Linear Loops

What is the simplest class of programs for which decidability of the Halting Problem is open?

$\mathbf{x} := \mathbf{a}$;
while $\mathbf{b} \cdot \mathbf{x} \neq 0$ do
    $\mathbf{x} := \mathbf{M} \cdot \mathbf{x}$;

Halting Problem for Single-Path Linear Loops

What is the simplest class of programs for which decidability of the Halting Problem is open?

$\mathbf{x} := \mathbf{a}$;
while $\mathbf{b} \cdot \mathbf{x} \neq 0$ do
    $\mathbf{x} := \mathbf{M} \cdot \mathbf{x}$;

Halting Problem

Instance: $\langle\ \mathbf{a}\ ;\ \mathbf{b}\ ;\ \mathbf{M}\ \rangle$
Question: Does this program halt?

Classical Formulation

A **linear recurrence sequence (LRS)** is a sequence $\langle u_0, u_1, u_2, \ldots \rangle$ in $\mathbb{Q}$ such that there are constants $a_1, \ldots, a_k$ and,

$$\forall n \geq 0 : \quad u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \ldots + a_k u_n.$$

Classical Formulation

A **linear recurrence sequence (LRS)** is a sequence $\langle u_0, u_1, u_2, \ldots \rangle$ in $\mathbb{Q}$ such that there are constants $a_1, \ldots, a_k$ and,

$$\forall n \geq 0 : \quad u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \ldots + a_k u_n.$$

- e.g. the Fibonacci numbers $\langle 0, 1, 1, 2, 3, 5, 8, \ldots \rangle$

Classical Formulation

A **linear recurrence sequence (LRS)** is a sequence $\langle u_0, u_1, u_2, \ldots \rangle$ in $\mathbb{Q}$ such that there are constants $a_1, \ldots, a_k$ and,

$$\forall n \geq 0 : \quad u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \ldots + a_k u_n.$$

- e.g. the Fibonacci numbers $\langle 0, 1, 1, 2, 3, 5, 8, \ldots \rangle$
- k is the **order** of the sequence
  - Fibonacci has order 2 $(u_{n+2} = u_{n+1} + u_n)$

Classical Formulation

A **linear recurrence sequence (LRS)** is a sequence $\langle u_0, u_1, u_2, \ldots \rangle$ in $\mathbb{Q}$ such that there are constants $a_1, \ldots, a_k$ and,
$\forall n \geq 0 : \quad u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \ldots + a_k u_n.$

- e.g. the Fibonacci numbers $\langle 0, 1, 1, 2, 3, 5, 8, \ldots \rangle$
- k is the **order** of the sequence
  - Fibonacci has order 2 $(u_{n+2} = u_{n+1} + u_n)$
- **Exponential-polynomial** closed form: $u_n = \sum_{i \in I} C_i(n) \lambda_i^n$:

Classical Formulation

A **linear recurrence sequence (LRS)** is a sequence $\langle u_0, u_1, u_2, \ldots \rangle$ in $\mathbb{Q}$ such that there are constants $a_1, \ldots, a_k$ and,
$\forall n \geq 0 : \quad u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \ldots + a_k u_n.$

- e.g. the Fibonacci numbers $\langle 0, 1, 1, 2, 3, 5, 8, \ldots \rangle$
- k is the **order** of the sequence
  - Fibonacci has order 2 $(u_{n+2} = u_{n+1} + u_n)$
- **Exponential-polynomial** closed form: $u_n = \sum_{i \in I} C_i(n) \lambda_i^n$:

Problem SKOLEM

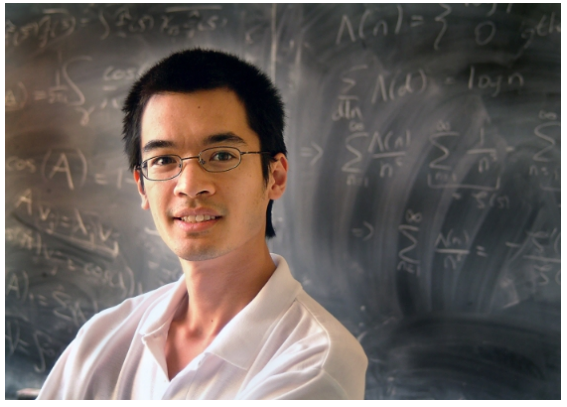
Instance: An LRS $\langle u_0, u_1, u_2, \ldots \rangle$
Question: Does $\exists n \geq 0$ such that $u_n = 0$?

Decidability is Open!

Decidability is Open!

"It is faintly outrageous that this problem is still open; it is saying that we do not know how to decide the Halting Problem even for 'linear' automata!"

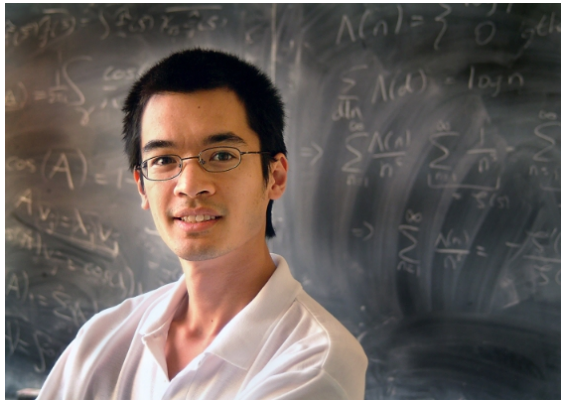
Terence Tao



Decidability is Open!

"It is faintly outrageous that this problem is still open; it is saying that we do not know how to decide the Halting Problem even for 'linear' automata!"

Terence Tao



"...a mathematical embarrassment..."

Richard Lipton

The Skolem-Mahler-Lech Theorem

Theorem (Skolem 1934; Mahler 1935, 1956; Lech 1953)

Let $\langle u_n \rangle$ be a **non-degenerate** *linear recurrence sequence that is not identically zero. Then the set $\{n : u_n = 0\}$ is finite.*

The Skolem-Mahler-Lech Theorem

Theorem (Skolem 1934; Mahler 1935, 1956; Lech 1953)

Let $\langle u_n \rangle$ be a **non-degenerate** *linear recurrence sequence that is not identically zero. Then the set $\{n : u_n = 0\}$ is finite.*

- Skolem's proof used p-adic analysis and is ineffective.

The Skolem-Mahler-Lech Theorem

Theorem (Skolem 1934; Mahler 1935, 1956; Lech 1953)

Let $\langle u_n \rangle$ be a **non-degenerate** *linear recurrence sequence that is not identically zero. Then the set $\{n : u_n = 0\}$ is finite.*

- Skolem's proof used p-adic analysis and is ineffective.
- From the early 2000s: series of papers that use the Subspace Theorem in Diophantine approximation to give explicit upper bounds on the number of zeros of a non-degenerate LRS.

The Skolem Problem at Low Orders

Theorem (folklore)

For orders 1 and 2, Skolem is decidable.

The Skolem Problem at Low Orders

Theorem (folklore)

For orders 1 and 2, Skolem is decidable.

Theorem (Mignotte, Shorey, Tijdeman 1984; Vereshchagin 1985)

For orders 3 and 4, Skolem is decidable.

The Skolem Problem at Low Orders

Theorem (folklore)

For orders 1 and 2, Skolem is decidable.

Theorem (Mignotte, Shorey, Tijdeman 1984; Vereshchagin 1985)

For orders 3 and 4, Skolem is decidable.

Critical ingredient is Baker's theorem for linear forms in logarithms, which earned Baker the Fields Medal in 1970.

The Skolem Problem at Low Orders

Theorem (folklore)

For orders 1 and 2, Skolem is decidable.

Theorem (Mignotte, Shorey, Tijdeman 1984; Vereshchagin 1985)

For orders 3 and 4, Skolem is decidable.

Critical ingredient is Baker's theorem for linear forms in logarithms, which earned Baker the Fields Medal in 1970.

Corollary

The Halting Problem is decidable for loops with at most 4 variables

Universal Skolem Set

Definition

Define $\mathcal{S} \subseteq \mathbb{N}$ to be a **Universal Skolem Set** if there is an effective procedure that inputs an integer linear recurrence sequence $\langle u_n \rangle$ and outputs whether or not there exists $n \in \mathcal{S}$ with $u_n = 0$.

Universal Skolem Set

Definition

Define $\mathcal{S} \subseteq \mathbb{N}$ to be a **Universal Skolem Set** if there is an effective procedure that inputs an integer linear recurrence sequence $\langle u_n \rangle$ and outputs whether or not there exists $n \in \mathcal{S}$ with $u_n = 0$.

Example

Define $f : \mathbb{N}_+ \to \mathbb{N}$ by $f(n) = \lfloor \sqrt{\log n} \rfloor$. Write $s_0 := 1$ and, inductively set $s_n := n! + s_{f(n)}$ for $n > 0$. Then $\mathcal{S} := \{s_n : n \in \mathbb{N}\}$ is a Universal Skolem Set.

Universal Skolem Set

Definition

Define $\mathcal{S} \subseteq \mathbb{N}$ to be a **Universal Skolem Set** if there is an effective procedure that inputs an integer linear recurrence sequence $\langle u_n \rangle$ and outputs whether or not there exists $n \in \mathcal{S}$ with $u_n = 0$.

Example

Define $f : \mathbb{N}_+ \to \mathbb{N}$ by $f(n) = \lfloor \sqrt{\log n} \rfloor$. Write $s_0 := 1$ and, inductively set $s_n := n! + s_{f(n)}$ for $n > 0$. Then $\mathcal{S} := \{s_n : n \in \mathbb{N}\}$ is a Universal Skolem Set.

- $\mathcal{S}$ has density zero:

$$|\mathcal{S} \cap \{1, \ldots, n\}| \sim \frac{\log n}{\log \log n}$$

A Universal Skolem Set of Positive Lower Density

- A **representation** of a positive integer n is a triple (P, q, a) such that $n = Pq + a$, P is prime, and q, a are $o(\log n)$.

A Universal Skolem Set of Positive Lower Density

- A **representation** of a positive integer n is a triple (P, q, a) such that $n = Pq + a$, P is prime, and q, a are $o(\log n)$.

- Define $\mathcal{U}$ to be the set of positive integers n with "many" (at least $\log_4 n$) representations.

A Universal Skolem Set of Positive Lower Density

- A **representation** of a positive integer n is a triple (P, q, a) such that $n = Pq + a$, P is prime, and q, a are $o(\log n)$.

- Define $\mathcal{U}$ to be the set of positive integers n with "many" (at least $\log_4 n$) representations.

Theorem

The set $\mathcal{U}$ is a Universal Skolem Set of positive lower density.

Simple Linear Recurrence Sequences

- An LRS is **simple** if its *characteristic roots* are simple

Simple Linear Recurrence Sequences

- An LRS is **simple** if its *characteristic roots* are simple
- Closed-form is a power sum $u_n = \sum_{i \in I} C_i \lambda_i^n$:

Simple Linear Recurrence Sequences

- An LRS is **simple** if its *characteristic roots* are simple
- Closed-form is a power sum $u_n = \sum_{i \in I} C_i \lambda_i^n$:

$$F_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$$

Simple Linear Recurrence Sequences

- An LRS is **simple** if its *characteristic roots* are simple
- Closed-form is a power sum $u_n = \sum_{i \in I} C_i \lambda_i^n$:

$$F_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$$

- Simple LRS correspond precisely to **diagonalisable** matrices

Exponential Diophantine Equations in Multiple Variables

Exponential Diophantine Equations in Multiple Variables

There are explicit upper bounds on number of solutions n of $\sum_{i \in I} C_i \lambda_i^n = 0$.

Exponential Diophantine Equations in Multiple Variables

There are explicit upper bounds on number of solutions n of $\sum_{i \in I} C_i \lambda_i^n = 0$.

Theorem (Schlickewei and Schmidt 2000)

Let α_i, β_i, C_i be non-zero algebraic numbers for $i = 1, \ldots, \ell$, Excepting some degenerate cases, the equation

$$\sum_{i=1}^{\ell} C_i \alpha_i^n \beta_i^m = 0 \tag{1}$$

has finitely many solutions in integers n, m. Moreover there is an explicit upper bound on the number of solutions in terms of ℓ and the degree of the α_i, β_i, C_i.

$\mathcal{U}$ is a Universal Skolem Set

- Consider non-degenerate **simple** LRS $u_n := \sum_{i \in I} C_i \lambda_i^n$:

$\mathcal{U}$ is a Universal Skolem Set

- Consider non-degenerate **simple** LRS $u_n := \sum_{i \in I} C_i \lambda_i^n$:

Proposition

If $u_n = 0$ and n has decomposition $n = qP + a$ then q, a solve a "companion equation".

$\mathcal{U}$ is a Universal Skolem Set

- Consider non-degenerate **simple** LRS $u_n := \sum_{i \in I} C_i \lambda_i^n$:

Proposition

If $u_n = 0$ and n has decomposition $n = qP + a$ then q, a solve a "companion equation".

- Previous result of Schlickewei–Schmidt induces upper bound on number of representations of n.

$\mathcal{U}$ is a Universal Skolem Set

- Consider non-degenerate **simple** LRS $u_n := \sum_{i \in I} C_i \lambda_i^n$:

Proposition

If $u_n = 0$ and n has decomposition $n = qP + a$ then q, a solve a "companion equation".

- Previous result of Schlickewei–Schmidt induces upper bound on number of representations of n.

Proposition

Let $\langle u_n \rangle$ be a non-degenerate simple LRS. Then there is a computable upper bound on the set $\{n \in \mathcal{U} : u_n = 0\}$.

On the Density of $\mathcal{U}$

Theorem

The set $\mathcal{U}$ has positive lower density.

On the Density of $\mathcal{U}$

Theorem

The set $\mathcal{U}$ has positive lower density.

Proof. Count pairs of representations $Pq + a = P'q' + a'$ that coincide.

On the Density of $\mathcal{U}$

Theorem

The set $\mathcal{U}$ has positive lower density.

Proof. Count pairs of representations $Pq + a = P'q' + a'$ that coincide.

Theorem (Sieve)

Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ be such that $|a_1 a_2(a_1 b_2 - a_2 b_1)| \neq 0$. Then

$$|\{t \leq X : a_1 t + b_1, a_2 t + b_2 \text{ both prime}\}| \ll \frac{X}{(\log X)^2}.$$

What is the Density of $\mathcal{U}$?

What is the Density of $\mathcal{U}$?

"In investigations concerning the asymptotic properties of arithmetic functions, it is often possible to make an interesting use of heuristic arguments."

Harald Cramér

What is the Density of $\mathcal{U}$?

"In investigations concerning the asymptotic properties of arithmetic functions, it is often possible to make an interesting use of heuristic arguments."

Harald Cramér

Cramér conjecture:

$$\max_{p_n \leq x}(p_{n+1} - p_n) = O(\log^2 x)$$

What is the Density of $\mathcal{U}$?

"In investigations concerning the asymptotic properties of arithmetic functions, it is often possible to make an interesting use of heuristic arguments."

Harald Cramér

Cramér conjecture:

$$\max_{p_n \leq x}(p_{n+1} - p_n) = O(\log^2 x)$$

"The exact formulation of Cramér's conjecture has been called into question. It is still probably true that for all $c > 2$,

$$\max_{p_n \leq x}(p_{n+1} - p_n) = O(\log^c x)\text{ "}$$

Adelman and McCurley

Heuristics for the Skolem Problem

Heuristics for the Skolem Problem

Theorem

There is a set $\mathcal{U}$ of positive lower density for which there is an algorithm that takes as input a non-degenerate LRS and outputs its set of zeros in $\mathcal{U}$.

Heuristics for the Skolem Problem

Theorem

There is a set $\mathcal{U}$ of positive lower density for which there is an algorithm that takes as input a non-degenerate LRS and outputs its set of zeros in $\mathcal{U}$.

- Cramér heuristic suggests that $\mathcal{U}$ has density one.