On Linear Recurrence Sequences and Loop Termination

Joël Ouaknine, Department of Computer Science, Oxford University, UK
James Worrell, Department of Computer Science, Oxford University, UK

A sequence of real numbers is said to be *positive* if all terms are positive, and *ultimately positive* if all but finitely many terms are positive. In this article we survey recent progress on long-standing open problems concerning deciding the positivity and ultimate positivity of integer linear recurrence sequences. We briefly discuss some of the many contexts in which these problems arise, and relate them to the well-known Skolem Problem, which asks whether a given linear recurrence sequence has a zero term. We also highlight some of the mathematical techniques that have been used to obtain decision procedures for these problems, pointing out obstacles to further progress.

In the second half of this survey we move on to closely related questions concerning the termination of linear while loops. This is a well-studied subject in software verification and by now there is rich toolkit of techniques to prove termination in practice. However the decidability of termination for some of the most basic types of loop remains open. Here again we discuss recent progress and remaining open problems.

1. INTRODUCTION

Linear recurrence sequences (LRS), such as the Fibonacci numbers, arise in a surprisingly diverse range of contexts, particularly in mathematics and computer science, but also in fields such as physics, biology, and economics. Connections to automata theory abound: for example, let A be a (deterministic or nondeterministic) finite automaton over some alphabet, and for every non-negative integer n, write u_n to denote the number of distinct words of length n accepted by A. Then the sequence $\langle u_0, u_1, u_2, \ldots \rangle$ of integers is an LRS. Moreover, given any integer LRS $\boldsymbol{v} = \langle v_0, v_1, v_2, \ldots \rangle$, one can always find two finite automata A and B such that each v_n is precisely the difference between the number of words of length n accepted by A and the number of words of length n accepted by B.

In this short survey, we examine connections between LRS and the termination of linear loops. The latter is a central problem in software verification that has attracted a substantial amount of attention over the last three decades, and has led to the development of tools such as Microsoft Research's TERMINATOR and T2 [Cook et al. 2011]. See also the surveys by Ben-Amram and Genaim [Ben-Amram and Genaim 2014], and by Gasarch [Gasarch 2015] describing semi-algorithmic approaches to termination based on *ranking functions*. By contrast, our interest here lies in *decidability* (and complexity) questions. Perhaps surprisingly, such questions are in turn related to deep problems and techniques from analytic and algebraic number theory, Diophantine geometry, and real algebraic geometry.

Let us start with some definitions. A (real) LRS is an infinite sequence $\boldsymbol{u} = \langle u_0, u_1, u_2, \ldots \rangle$ of real numbers having the following property: there exist real constants $a_1, a_2, \ldots, a_k$ such that, for all $n \geq 0$,

$$u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \ldots + a_k u_n \, . \tag{1}$$

If the initial values $u_0, \ldots, u_{k-1}$ of the sequence are provided, the recurrence relation defines the rest of the sequence uniquely. The smallest k for which an LRS obeys such a recurrence relation is the ***order*** of the LRS.

Given an LRS $\boldsymbol{u}$ satisfying the recurrence relation (1), the ***characteristic polynomial*** of $\boldsymbol{u}$ is

$$p(x) = x^k - a_1 x^{k-1} - \ldots - a_{k-1} x - a_k \, .$$

An LRS is said to be ***simple*** if its characteristic polynomial has no repeated roots. Simple LRS, such as the Fibonacci sequence, possess a number of desirable properties which considerably simplify their analysis. They constitute a large and well-studied class of sequences, and correspond to the iterated application of *diagonalisable* matrices, as we explain shortly.

Real LRS can equivalently be characterised as those sequences $\boldsymbol{u} = \langle u_0, u_1, u_2, \ldots \rangle$ admitting an *exponential-polynomial* representation

$$u_n = \sum_{i=1}^{r} P_i(n)\rho_i^n + \sum_{i=1}^{s} Q_i(n)\lambda_i^n + \overline{Q_i(n)\lambda_i^n}\,,$$

where the ρ_i are distinct real numbers, the P_i are non-zero polynomials with real coefficients, the λ_i are distinct complex numbers, and the Q_i are non-zero polynomials with complex coefficients. Such a sequence is an LRS of order $r+2s$, with characteristic roots $\rho_1, \ldots, \rho_r, \lambda_1, \overline{\lambda_1}, \ldots, \lambda_s, \overline{\lambda_s}$. Under this representation, $\boldsymbol{u}$ is a simple LRS if and only if each of the polynomial terms is constant.

Motivated by questions in language theory and formal power series, Rozenberg, Salomaa, and Soittola [Salomaa and Soittola 1978; Rozenberg and Salomaa 1994] highlighted the following four decision problems concerning LRS over rational numbers. Given an LRS $\boldsymbol{u}$ as above:

(1) Does $u_n = 0$ for some n?
(2) Does $u_n = 0$ for infinitely many n?
(3) Is $u_n \geq 0$ for all n?
(4) Is $u_n \geq 0$ for all but finitely many n?

Problem 1 is known as the ***Skolem Problem***, after the Skolem-Mahler-Lech Theorem [Everest et al. 2003; Halava et al. 2005] which characterises the set $\{n \in \mathbb{N} : u_n = 0\}$ of zeros of an LRS $\boldsymbol{u}$ as semi-linear, i.e., consisting of a finite set together with a finite number of (infinite) arithmetic progressions. The proof of the latter is however non-effective, and the decidability of the Skolem Problem is generally considered to have been open for over 80 years [Tao 2008], a state of affairs described as "faintly outrageous" by Tao [Tao 2008] and a "mathematical embarrassment" by Lipton [Lipton and Regan 2013]. A breakthrough occurred in the mid-1980s, when Mignotte *et al.* [Mignotte et al. 1984] and Vereshchagin [Vereshchagin 1985] independently showed decidability for LRS of order 4 or less. These deep results make essential use of Baker's theorem on linear forms in logarithms (which earned Baker the Fields Medal in 1970), as well as a p-adic analogue of Baker's theorem due to van der Poorten. Unfortunately, little progress on that front has since been recorded.[1] The Skolem Problem can also be seen as a generalisation of the Orbit Problem, studied by Kannan and Lipton [Kannan and Lipton 1986, Sec. 5].

Interestingly and in contrast, Problem 2—hitting zero infinitely often—was shown to be decidable for arbitrary LRS by Berstel and Mignotte [Berstel and Mignotte 1976].

Problems 3 and 4 are respectively known as the ***Positivity*** and the ***Ultimate Positivity*** Problems. It is considered folklore that the decidability of Positivity (for arbitrary LRS) would entail that of the Skolem Problem [Ouaknine and Worrell 2014b], noting however that the reduction increases the order of LRS quadratically. Both Positivity and Ultimate Positivity are stated as open in literature going back at least to the 1970s (see, e.g., [Soittola 1976; Salomaa 1976; Berstel and Mignotte 1976]), as well as more recently (cf. [Halava et al. 2006; Bell and Gerhold 2007; Laohakosol and Tangsupphathawat 2009; Liu 2010; Tarasov and Vyalyi 2011; Ouaknine and Worrell 2014b], among others).

[1] A proof of decidability of the Skolem Problem for LRS of order 5 was announced in [Halava et al. 2005]. However, as pointed out in [Ouaknine and Worrell 2012], the proof seems to have a serious gap.

Hitherto, all decidability results for Positivity and Ultimate Positivity have been for low-order LRS; the paper [Ouaknine and Worrell 2014b] gives a detailed account of these results, obtained over a period of time stretching back some 35 years, and proves decidability of both problems for sequences of order at most 5 (with complexity in the Counting Hierarchy, itself contained within PSPACE). In addition, it is shown in [Ouaknine and Worrell 2014b] that obtaining decidability for either Positivity or Ultimate Positivity at order 6 would necessarily entail major breakthroughs in analytic number theory (more precisely regarding long-standing open problems in Diophantine approximation of transcendental numbers).

For *simple* LRS, Positivity is known to be decidable up to order 9 [Ouaknine and Worrell 2014a], and Ultimate Positivity is decidable for all orders [Ouaknine and Worrell 2014c] (in PSPACE when the order is not fixed, with a nearly matching lower bound, and in P for any fixed order—see [Ouaknine and Worrell 2014c] for details). However, beyond order 9, the algorithm for Ultimate Positivity is *non-constructive*: given an ultimately periodic LRS $\langle u_n \rangle_{n=0}^{\infty}$, the procedure of [Ouaknine and Worrell 2014c] does not produce a threshold N such that $u_n \geq 0$ for all $n \geq N$; indeed the ability to compute such a threshold would immediately yield an algorithm for the Positivity Problem for simple LRS of all orders, since the signs of $u_0, \ldots, u_{N-1}$ can be evaluated directly. In turn this would yield decidability of the Skolem Problem for simple LRS (known to be open) and also would enable one to decide the Skolem Problem for general LRS of order 5, as the only remaining unsolved sub-case involves distinct characteristic roots (cf. [Ouaknine and Worrell 2012]).

The non-constructive nature of the algorithm for deciding Ultimate Positivity of simple LRS arises from the use of lower bounds in Diophantine approximation concerning *sums of S-units*. These bounds were established in [Evertse 1984; van der Poorten and Schlickewei 1982] using Schlickewei's p-adic generalisation of Schmidt's Subspace Theorem (itself a far-reaching generalisation of the Thue-Siegel-Roth Theorem), and therein applied to study the asymptotic growth of LRS in absolute value. By contrast, [Ouaknine and Worrell 2014a] invokes Baker's Theorem on linear forms in logarithms to show decidability of Positivity for simple LRS of order at most 9. Unfortunately, while Baker's Theorem yields effective Diophantine-approximation lower bounds, it appears only to be applicable to low-order LRS. In particular, the analytic and geometric arguments that are used in [Ouaknine and Worrell 2014a] to bring Baker's Theorem to bear do not seem applicable beyond order 9.

2. EXTENDED EXAMPLE

The following extended example illustrates some of the main ideas involved in deciding positivity of an LRS, as well as some of the obstacles to generalising the decidability results described in the previous section. Full details of the method illustrated here can be found in [Ouaknine and Worrell 2014a].

Consider the sequence $\boldsymbol{u}$ given by the exponential-polynomial expression

$$u_n = \frac{33}{8} + \lambda_1^n + \overline{\lambda_1^n} + 2\lambda_2^n + 2\overline{\lambda_2^n}\,,$$

where

$$\lambda_1 = \frac{-3+4i}{5} \quad \text{and} \quad \lambda_2 = \frac{-7+24i}{25}\,.$$

This expression defines a sequence of rational numbers satisfying an order-5 linear recurrence with rational coefficients:

$$u_{n+5} = -\frac{19}{25}u_{n+4} - \frac{114}{125}u_{n+3} + \frac{114}{125}u_{n+2} + \frac{19}{25}u_{n+1} + u_n$$

whose characteristic roots are 1, λ_1, λ_2, $\overline{\lambda_1}$, and $\overline{\lambda_2}$. In this example we consider how to establish the positivity of $\boldsymbol{u}$, that is, whether $u_n \geq 0$ for all $n \in \mathbb{N}$.

Notice first that λ_1 and λ_2 both lie on the unit circle $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ in the complex plane. Moreover, neither λ_1 nor λ_2 is a complex root of unity, so their respective orbits $\{\lambda_1^n : n \in \mathbb{N}\}$ and $\{\lambda_2^n : n \in \mathbb{N}\}$ are both dense in $\mathbb{T}$. In light of this one might be led to believe that u_n can be negative, e.g., if n is such that $\mathrm{Re}(\lambda_1^n)$ and $\mathrm{Re}(\lambda_2^n)$ are both at most $-\frac{3}{4}$ then clearly $u_n < 0$. Crucially, however, the joint orbit $\{(\lambda_1^n, \lambda_2^n) : n \in \mathbb{N}\}$ *is not* dense in $\mathbb{T}^2$. This is because λ_1 and λ_2 satisfy the multiplicative relationship $\lambda_1^2\lambda_2 = 1$. From this it is immediate that $\{(\lambda_1^n, \lambda_2^n) : n \in \mathbb{N}\}$ is contained in the set

$$\boldsymbol{T} := \{(z_1, z_2) \in \mathbb{T}^2 : z_1^2 z_2 = 1\} .$$

In fact, the set of all multiplicative relationships

$$\{(n_1, n_2) \in \mathbb{Z}^2 : \lambda_1^{n_1}\lambda_2^{n_2} = 1\}$$

among λ_1 and λ_2 is a rank-1 subgroup of $\mathbb{Z}^2$ with generator $(2, 1)$. It follows from Kronecker's theorem on simultaneous inhomogeneous Diophantine approximation [Cassels 1965] that $\{(\lambda_1^n, \lambda_2^n) : n \in \mathbb{N}\}$ is a dense subset of $\boldsymbol{T}$.

Now consider the function $f : \mathbb{T}^2 \to \mathbb{R}$ given by

$$f(z_1, z_2) = \frac{33}{8} + z_1 + \overline{z_1} + 2z_2 + 2\overline{z_2} .$$

Then $u_n = f(\lambda_1^n, \lambda_2^n)$ and (by calculus) f is non-negative on the set $\boldsymbol{T}$. It follows that $u_n \geq 0$ for all n.[2]

Next we extend the example by considering the order-6 LRS $\boldsymbol{v} = \langle v_0, v_1, v_2, \ldots\rangle$, given by $v_n := u_n - \frac{1}{2^n}$. We have already established positivity of $\boldsymbol{u}$. Positivity of $\boldsymbol{v}$ amounts to establishing the lower bound $\frac{1}{2^n} \leq u_n$ for all $n \in \mathbb{N}$. But this is a delicate question because u_n comes arbitrarily close to 0 as n ranges over $\mathbb{N}$. Indeed the function f above has two zeros on the set $\boldsymbol{T}$—at the point

$$(z_1^*, z_2^*) := \left(-\tfrac{1}{8} + \tfrac{\sqrt{63}}{8}i, -\tfrac{31}{32} + \tfrac{\sqrt{63}}{32}i\right)$$

and its complex conjugate. Since $\{(\lambda_1^n, \lambda_2^n) : n \in \mathbb{N}\}$ is a dense subset of $\boldsymbol{T}$ it follows that $\liminf_n u_n = \liminf_n f(\lambda_1^n, \lambda_2^n) = 0$.

One can obtain a lower bound for $f(\lambda_1^n, \lambda_2^n)$ as a function of n via lower bounds on the distance between $(\lambda_1^n, \lambda_2^n)$ and the two zeros of f in $\boldsymbol{T}$. Baker's Theorem on linear forms in logarithms of algebraic numbers is instrumental in obtaining such bounds.

A simple form of Baker's Theorem (restricted to linear forms in three logarithms) is as follows. Let $\alpha_1, \alpha_2, \alpha_3$ be algebraic numbers and b_1, b_2, b_3 integers of absolute value at most H. Then for the principal branch $\log$ of the complex logarithm function,

$$\Lambda := b_1 \log \alpha_1 + b_2 \log \alpha_2 + b_3 \log \alpha_3$$

is either 0 or satisfies $|\Lambda| > H^{-C}$, where C is an effectively computable constant depending only on α_1, α_2, and α_3.

With this result in hand, we can bound the distance between λ_1^n and z_1^* as follows. Note that for any $\alpha \in \mathbb{T}$, we have $\log \alpha = i \arg \alpha$. Next, let n be such that $\lambda_1^n \neq z_1^*$ and

[2] Note that density of $\{(\lambda_1^n, \lambda_2^n) : n \in \mathbb{N}\}$ in $\boldsymbol{T}$ is not needed to certify the positivity of $\boldsymbol{u}$, only inclusion. However the fact that the topological closure of the orbit $\{(\lambda_1^n, \lambda_2^n) : n \in \mathbb{N}\}$ is an algebraic subset of $\mathbb{C}^2$ is an important property that is relevant to the completeness of method of proving positivity illustrated in this example.

choose $m \in \mathbb{Z}$ such that $-\pi < n \arg \lambda_1 - 2\pi m \leq \pi$. Then

$$\begin{aligned} |\lambda_1^n - z_1^*| &\geq \frac{1}{2} \, |n \arg \lambda_1 - 2\pi m - \arg z_1^*| \\ &\geq \frac{1}{\operatorname{poly}(n)} \, , \end{aligned} \tag{2}$$

where $\operatorname{poly}(n)$ is an effectively computable polynomial depending only on λ_1 and z_1^*. Here the first inequality involves estimating the distance between λ_1^n and z_1^* in terms of the length of the circular segment of $\mathbb{T}$ lying between λ_1^n and z_1^*. The second inequality follows from Baker's Theorem, using the facts that $\pi = \arg(-1)$ and $|m|$ is less than n.

We can now compute $N \in \mathbb{N}$ such that $u_n \geq \frac{1}{2^n}$ for all $n \geq N$. First, invoking ideas similar to those used by Kannan and Lipton in their work on the Orbit Problem [Kannan and Lipton 1986], one can choose N sufficiently large such that $(\lambda_1^n, \lambda_2^n)$ is not equal to either or the two zeros of the function f on $\boldsymbol{T}$ for $n \geq N$. One can then apply inequalities of type (2) to get lower bounds for the distance between $(\lambda_1^n, \lambda_2^n)$ and the zeros of f, and hence obtain a lower bound on $u_n = f(\lambda_1^n, \lambda_2^n)$ of the form $\frac{1}{\operatorname{poly}(n)}$ for some fixed effectively computable polynomial $\operatorname{poly}(n)$. Choosing N such that $\frac{1}{\operatorname{poly}(n)} \geq \frac{1}{2^n}$ for all $n \geq N$, we have $u_n \geq \frac{1}{2^n}$ for all $n \geq N$. It immediately follows that $v_n < 0$ can only happen if $n < N$. Thus the positivity of $\boldsymbol{v}$ can be established by checking positivity of v_n for $n < N$ by exhaustive search.

We conclude the example with several remarks. First, recall that an important step in establishing the positivity of $\boldsymbol{u}$ was to show that the function f only assumes non-negative values on $\boldsymbol{T}$. In the case at hand, this fact could be shown using elementary calculus. More generally one can phrase such a question as the truth of a universal sentence in the theory of real-closed fields. This is the approach taken in [Ouaknine and Worrell 2014c], which gives a procedure to decide the ultimate positivity of simple linear recurrence sequences. In the other direction, we have also shown in [Ouaknine and Worrell 2014c] that deciding the truth of such universal sentences is reducible in polynomial time to deciding the ultimate positivity of simple linear recurrence sequences.

Next, observe that our analysis of the positivity of $\boldsymbol{v}$ crucially depends on computing an *effective* sub-exponential lower bound on u_n. We were able to do this thanks to the effective nature of the constants in Baker's Theorem. In turn, our application of Baker's Theorem hinged on the fact that the function f had only isolated zeros on $\boldsymbol{T}$. This isolated-zero condition can be shown to fail for higher-order recurrences. In particular, as we have described in the previous section, our result that ultimate positivity is decidable for simple LRS of arbitrary order relies on non-effective Diophantine-approximation lower bounds in place of Baker's Theorem [Ouaknine and Worrell 2014c]. Such results suffice for considering *ultimate* positivity as well as for handling termination of linear while loops, as we will see in the next section.

Finally, observe that the sequence $\boldsymbol{v}$ in the above example is simple. Even at order 6, if one admits linear terms in exponential polynomials then to decide positivity one provably needs much sharper Diophantine-approximation estimates for logarithms of algebraic numbers than are currently known; see [Ouaknine and Worrell 2014b] for details.

3. TERMINATION OF LINEAR LOOPS

Termination is a fundamental decision problem in program verification. In particular, termination of programs with linear assignments and guards has been extensively studied over the last decade; as discussed earlier, this has led to the development of powerful semi-algorithms to prove termination via synthesis of ranking functions, several of which have been implemented in software-verification tools.

In this survey, we focus on ***linear loops***, i.e., programs of the form:

$$\mathsf{P}: \ \boldsymbol{x} \leftarrow \boldsymbol{t}\,; \textit{while } B\boldsymbol{x} \geq \boldsymbol{c} \textit{ do } \boldsymbol{x} \leftarrow A\boldsymbol{x} + \boldsymbol{a}\,,$$

where $\boldsymbol{x}$ is vector of variables, $\boldsymbol{t}$ is a vector of integer, rational, or real numbers, $\boldsymbol{a}$ and $\boldsymbol{c}$ are integer vectors, and A and B are integer matrices of the appropriate dimensions. Here the loop guard is a conjunction of linear inequalities and the loop body consists of a simultaneous affine assignment to $\boldsymbol{x}$. When vectors $\boldsymbol{a}$ and $\boldsymbol{c}$ are both zero, the loop is said to be ***homogeneous***.

The ***dimension*** of a linear loop is that of vector $\boldsymbol{x}$. For P of dimension d, we say that P ***terminates*** on a set $S \subseteq \mathbb{R}^d$ if it terminates for all initial vectors $\boldsymbol{t} \in S$. In 2004, Tiwari gave a procedure to decide whether a given linear loop terminates on $\mathbb{R}^d$ [Tiwari 2004], and two years later Braverman showed decidability of termination on $\mathbb{Q}^d$ [Braverman 2006]. However the most natural problems from the point of view of program verification are (i) termination on $\mathbb{Z}^d$ and (ii) termination on given integer singleton sets (corresponding to fixed initial starting conditions, and also known as the *Halting Problem* for linear loops).

While termination on $\mathbb{Z}^d$ reduces to termination on $\mathbb{Q}^d$ in the homogeneous case (by a straightforward scaling argument), termination on $\mathbb{Z}^d$ in the general case is stated as an open problem in [Ben-Amram et al. 2012; Braverman 2006; Tiwari 2004]. Recently, we established decidability of termination on $\mathbb{Z}^d$ for linear loops provided either that such loops have dimension at most 4 or that the update matrix A is diagonalisable [Ouaknine et al. 2015]. (The general case however remains open.)

Two observations are in order. First—recalling that the guard of linear loops consists of a conjunction of linear inequalities—a given linear loop eventually halts if and only if one of the linear inequalities is eventually violated. In the study of linear loop termination, it is therefore sufficient to restrict one's attention to loops with a single inequality as guard. Second, given an inhomogeneous linear loop P of dimension d as above, one can readily manufacture a *homogeneous* linear loop P' such that P halts on a given initial starting vector $\boldsymbol{t} \in \mathbb{R}^d$ if and only if P' halts on $\boldsymbol{t}, 1 \in \mathbb{R}^{d+1}$ (i.e., the vector $\boldsymbol{t}$ augmented by a $(d+1)$th entry of 1. When studying termination on singleton sets, we may therefore assume homogeneity at the cost of increasing the dimension by 1. Note however that the attendant transformation may fail to preserve certain properties of the update matrix, such as diagonalisability.

Let us therefore examine the following d-dimensional homogeneous linear loop:

$$\mathsf{Q}: \ \boldsymbol{x} \leftarrow \boldsymbol{t}\,; \textit{while } \boldsymbol{b}^T\boldsymbol{x} \geq 0 \textit{ do } \boldsymbol{x} \leftarrow A\boldsymbol{x}\,,$$

where $\boldsymbol{b}^T$ is a row vector. Writing $u_n = \boldsymbol{b}^T A^n \boldsymbol{t}$, easily follows from the Cayley-Hamilton theorem that $\boldsymbol{u} = \langle u_0, u_1, u_2, \ldots \rangle$ is an LRS of order at most d. (Conversely, any LRS of order d may be realised as such a d-dimensional linear loop.) One can moreover show that $\boldsymbol{u}$ is simple if and only if A is diagonalisable.

From the above observations and our earlier results on LRS, we immediately obtain the following: it is decidable, for any fixed starting vector $\boldsymbol{t} \in \mathbb{Z}^d$, whether Q terminates provided either that d is at most 5, or that A is diagonalisable and d is at most 9. Moreover, by way of hardness, decidability of termination in dimension 6 or above (in the non-diagonalisable case) would necessarily entail major breakthroughs in analytic number theory. Turning to the inhomogeneous program P, one can decide termination on a singleton set provided that P has dimension 4 or less, simply by homogenising.

Let us return to the question of the termination of P on $\mathbb{Z}^d$, under the assumption that A is diagonalisable. This problem can equivalently be posed in terms of whether the set NT of initial values $\boldsymbol{t} \in \mathbb{R}^d$ on which P is non-terminating contains an integer point. The set NT is easily seen to be convex. In the following example it also happens to be a *semi-algebraic* subset of $\mathbb{R}^d$, i.e., defined by a conjunction of polynomial inequalities with integer coefficients.

Example 3.1. Let θ be a fixed real number that is not a rational multiple of π, and consider the program

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \leftarrow \boldsymbol{t}\,; \textit{\textbf{while}}\ z - y \geq 0\ \textit{\textbf{do}}\ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \leftarrow \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The update matrix in the loop body is a counter-clockwise rotation around the z-axis by angle θ. The set of non-terminating points is thus the cone $\{(x, y, z) \in \mathbb{R}^3 : z^2 \geq x^2 + y^2\}$.

Note that the update matrix in Example 3.1 is both diagonalisable and orthogonal. It follows that it has a basis of eigenvectors, all corresponding to eigenvalues of modulus one. More generally, consider an inhomogeneous linear loop with a diagonalisable update matrix A. Writing $\mathbb{R}^d$ as a direct sum $\mathbb{R}^d = V_1 \oplus \ldots \oplus V_m$, where each V_i is a sum of eigenspaces of A corresponding to eigenvalues of a given fixed modulus, it is shown in [Ouaknine et al. 2015] that for each i the set $\mathsf{NT} \cap V_i$ is an (effectively computable) semi-algebraic subset of $\mathbb{R}^d$. A key step to proving this result is to show how to compute all integer linear relationships among the phases of the eigenvalues associated to each V_i. Such a computation is possible thanks to number-theoretic results of Masser [Masser 1988].

While we know that $\mathsf{NT} \cap V_i$ is semi-algebraic for $i = 1, \ldots, m$, whether the set NT itself is semi-algebraic appears to be a much harder question. However for the purposes of deciding termination, this issue can be side-stepped by focusing on *eventual non-termination* rather than non-termination. We say that P is eventually non-terminating on $\boldsymbol{t} \in \mathbb{R}^d$ if, starting from initial value $\boldsymbol{t}$, after executing the loop body $\boldsymbol{x} \leftarrow A\boldsymbol{x} + \boldsymbol{a}$ a finite number of times *whilst disregarding the loop guard*, we eventually reach a value on which P fails to terminate. We write ENT for the set of eventually non-terminating initial vectors. Clearly and $\mathsf{NT} \subseteq \mathsf{ENT}$ and P is non-terminating if and only if ENT contains an integer point.

Decidability of termination in arbitrary dimension (assuming a diagonalisable update matrix) was shown in [Ouaknine et al. 2015] through an analysis of the set ENT. Given a linear loop, it is shown in [Ouaknine et al. 2015] how to compute a convex semi-algebraic set $W \subseteq \mathbb{R}^d$ such that the integer points $\boldsymbol{t} \in W$ are precisely the eventually non-terminating integer initial values. Since, by a result of Khachiyan and Porkolab [Khachiyan and Porkolab 1997], it is decidable whether a convex semi-algebraic set contains an integer point [Khachiyan and Porkolab 1997],[3] one can decide whether a linear loop is terminating on $\mathbb{Z}^d$. The computation of W makes critical use of deep number-theoretic tools such as the S-units theorem of Evertse, van der Poorten, and Schlickewei [Evertse 1984; van der Poorten and Schlickewei 1982], which as mentioned earlier played a key role in establishing decidability of Ultimate Positivity for simple LRS of all orders. Roughly speaking, these tools are used to prove that the termination of P on a given initial vector $\boldsymbol{t}$ is determined by the components of $\boldsymbol{t}$ on eigenspaces corresponding to eigenvalues of maximum modulus. Critically the non-effectiveness of these results is not a problem when considering eventual non-termination.

[3] By contrast, recall that the existence of an integer point in an *arbitrary* (i.e., not necessarily convex) semi-algebraic set—which is equivalent to Hilbert's tenth problem—is well-known to be undecidable.

The set W is used in [Ouaknine et al. 2015] in lieu of the set ENT of eventually non-terminating points. It is immediate from the definition of W that W and ENT have the same topological closure, and thus [Ouaknine et al. 2015] shows *inter alia* that the topological closure of ENT is semi-algebraic. Whether ENT or NT are semi-algebraic is, to the best of our knowledge, open. We refer the reader to [Ouaknine et al. 2015] for full details.

With regard to complexity, termination of linear loops over the set of all integer points is easily seen to be coNP-hard, by reduction from integer programming, taking the update matrix A to be the identity. The procedure proposed in [Ouaknine et al. 2015], on the other hand, requires exponential space. In contrast, even though not stated explicitly in [Tiwari 2004] and [Braverman 2006], deciding termination on $\mathbb{R}^d$ and $\mathbb{Q}^d$ (which relies mainly on spectral techniques and linear algebra) can be done in polynomial time.[4]

4. OPEN PROBLEMS

Decidability of termination of linear loops remains open if we do not assume a diagonalisable update matrix. Decidability of termination is also open for the more general class of ***linear constraint loops***, that is, loops of the form

$$\boldsymbol{x} \leftarrow \boldsymbol{t}\,;\, \textit{while}\ B\boldsymbol{x} \geq \boldsymbol{c}\ \textit{do}\ A\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{x}'\end{pmatrix} \leq \boldsymbol{d}\ ,$$

in which the loop body consists of a conjunction of linear constraints among the "before" and "after" values of the program variables, respectively denoted $\boldsymbol{x}$ and $\boldsymbol{x}'$.

A special case of linear constraint loops feature ***octagonal constraints***, which have the form $x - y' \sim k$ or $x + y' \sim k$ for (possibly identical) program variables x and y, integer k, and comparison operator $\sim \in \{\leq, \geq\}$. It is shown in [Bozga et al. 2012] that for linear constraint loops with exclusively octagonal constraints, the set of integer points in NT is effectively semi-linear. In fact more is true: for such loops the reflexive transitive closure of the transition relation denoted by the loop body is an effectively computable semi-linear set.

The outstanding open problem in the area is the decidability of Skolem's Problem. At we have remarked above, currently decidability is only known for LRS of order at most 4. While decidability at order 2 is elementary, the proofs of decidability at orders 3 and 4 make use of Baker's Theorem. On the other hand, unlike for the Positivity Problem, it is not clear that a solution to Skolem's Problem necessarily entails progress in understanding Diophantine approximation for logarithms of algebraic numbers. In terms of computational complexity, the current best lower bound is NP-hardness (which already holds for the restricted case of matrices with all entries either 0 or 1).

REFERENCES

J. P. Bell and S. Gerhold. 2007. On the Positivity Set of a Linear Recurrence. *Israel Jour. Math.* 57 (2007).

A. M. Ben-Amram and S. Genaim. 2014. Ranking Functions for Linear-Constraint Loops. *J. ACM* 61, 4 (2014).

A. M. Ben-Amram, S. Genaim, and A. N. Masud. 2012. On the Termination of Integer Loops. *ACM Trans. Program. Lang. Syst.* 34, 4 (2012).

J. Berstel and M. Mignotte. 1976. Deux propriétés décidables des suites récurrentes linéaires. *Bull. Soc. Math. France* 104 (1976).

M. Bozga, R. Iosif, and F. Konecný. 2012. Deciding conditional termination. In *Proc. TACAS (LNCS)*, Vol. 7214. Springer.

[4] This observation relies on the facts that one can compute Jordan canonical forms of integer matrices and solve instances of linear programming problems with algebraic numbers in polynomial time [Cai 1994].

M. Braverman. 2006. Termination of integer linear programs. In *Proc. CAV (LNCS)*, Vol. 4144. Springer.

J.-Y. Cai. 1994. Computing Jordan Normal Forms Exactly for Commuting Matrices in Polynomial Time. *Int. J. Found. Comput. Sci.* 5, 3/4 (1994), 293–302.

J.W.S. Cassels. 1965. *An introduction to Diophantine approximation*. Camb. Univ. Press.

B. Cook, A. Podelski, and A. Rybalchenko. 2011. Proving Program Termination. *Commun. ACM* 54, 5 (2011).

G. Everest, A. van der Poorten, I. Shparlinski, and T. Ward. 2003. *Recurrence Sequences*. American Mathematical Society.

J.-H. Evertse. 1984. On sums of S-units and linear recurrences. *Compositio Mathematica* 53, 2 (1984).

W. I. Gasarch. 2015. Proving Programs Terminate Using Well-Founded Orderings, Ramsey's Theorem, and Matrices. *Advances in Computers* 97 (2015), 147–200.

V. Halava, T. Harju, and M. Hirvensalo. 2006. Positivity of second order linear recurrent sequences. *Discrete Applied Mathematics* 154, 3 (2006).

V. Halava, T. Harju, M. Hirvensalo, and J. Karhumäki. 2005. *Skolem's Problem – On the Border between Decidability and Undecidability*. Technical Report 683. Turku Centre for Computer Science.

R. Kannan and R. J. Lipton. 1986. Polynomial-Time Algorithm for the Orbit Problem. *J. ACM* 33, 4 (1986).

L. Khachiyan and L. Porkolab. 1997. Computing Integral Points in Convex Semi-algebraic Sets. In *Proc. FOCS*. 162–171.

V. Laohakosol and P. Tangsupphathawat. 2009. Positivity of third order linear recurrence sequences. *Discrete Applied Mathematics* 157, 15 (2009).

R. J. Lipton and K. W. Regan. 2013. *People, Problems, and Proofs - Essays from Gödel's Lost Letter: 2010*. Springer.

L. Liu. 2010. Positivity of Three-Term Recurrence Sequences. *Elec. J. Comb.* 17, 1 (2010).

D. W. Masser. 1988. Linear Relations on Algebraic Groups. In *New Advances in Transcendence Theory*. Camb. Univ. Press.

M. Mignotte, T. Shorey, and R. Tijdeman. 1984. The distance between terms of an algebraic recurrence sequence. *J. für die reine und angewandte Math.* 349 (1984).

J. Ouaknine, J. Sousa Pinto, and J. Worrell. 2015. On Termination of Integer Linear Loops. In *Proc. SODA*. ACM-SIAM.

J. Ouaknine and J. Worrell. 2012. Decision Problems for Linear Recurrence Sequences. In *Proc. RP (LNCS)*, Vol. 7550. Springer.

J. Ouaknine and J. Worrell. 2014a. On the Positivity Problem for Simple Linear Recurrence Sequences. In *Proc. ICALP (LNCS)*, Vol. 8573. Springer.

J. Ouaknine and J. Worrell. 2014b. Positivity Problems for Low-Order Linear Recurrence Sequences. In *Proc. SODA*. ACM-SIAM.

J. Ouaknine and J. Worrell. 2014c. Ultimate Positivity is Decidable for Simple Linear Recurrence Sequences. In *Proc. ICALP (LNCS)*, Vol. 8573. Springer.

G. Rozenberg and A. Salomaa. 1994. *Cornerstones of Undecidability*. Prentice Hall.

A. Salomaa. 1976. Growth Functions of Lindenmayer Systems: Some New Approaches. In *Automata, Languages, Development*, A. Lindenmayer and G. Rozenberg (Eds.). North-Holland.

A. Salomaa and M. Soittola. 1978. *Automata-theoretic aspects of formal power series.* Springer.

M. Soittola. 1976. On D0L Synthesis Problem. In *Automata, Languages, Development*, A. Lindenmayer and G. Rozenberg (Eds.). North-Holland.

T. Tao. 2008. *Structure and randomness: pages from year one of a mathematical blog*. American Mathematical Society.

S. P. Tarasov and M. N. Vyalyi. 2011. Orbits of Linear Maps and Regular Languages. In *Proc. CSR (LNCS)*, Vol. 6651. Springer.

A. Tiwari. 2004. Termination of Linear Programs. In *Proc. CAV (LNCS)*, Vol. 3114. Springer.

A.J. van der Poorten and H.P. Schlickewei. 1982. The growth conditions for recurrence sequences. *Macquarie Math. Reports* 82-0041 (1982).

N. K. Vereshchagin. 1985. The problem of appearance of a zero in a linear recurrence sequence (in Russian). *Mat. Zametki* 38, 2 (1985).