What's decidable about linear loops?

Toghrul Karimov, Engel Lefaucheux, Joel Ouaknine, David Purser, Anton Varonka, Markus A. Whiteland, James Worrell

What's Decidable about Linear Loops?

TOGHRUL KARIMOV, Max Planck Institute for Software Systems, Germany
ENGEL LEFAUCHEUX, Max Planck Institute for Software Systems, Germany
JOËL OUAKNINE, Max Planck Institute for Software Systems, Germany
DAVID PURSER, Max Planck Institute for Software Systems, Germany
ANTON VARONKA, Max Planck Institute for Software Systems, Germany and Saarbrücken Graduate School of Computer Science, Germany
MARKUS A. WHITELAND, Max Planck Institute for Software Systems, Germany
JAMES WORRELL, University of Oxford, UK

We consider the MSO model-checking problem for simple linear loops, or equivalently discrete-time linear dynamical systems, with semialgebraic predicates (i.e., Boolean combinations of polynomial inequalities on the variables). We place no restrictions on the number of program variables, or equivalently the ambient dimension. We establish decidability of the model-checking problem provided that each semialgebraic predicate *either* has intrinsic dimension at most 1, *or* is contained within some three-dimensional subspace. We also note that lifting either of these restrictions and retaining decidability would necessarily require major breakthroughs in number theory.

CCS Concepts: • **Theory of computation → Logic and verification**.

Additional Key Words and Phrases: linear dynamical systems, MSO model checking, verification

ACM Reference Format:
Toghrul Karimov, Engel Lefaucheux, Joël Ouaknine, David Purser, Anton Varonka, Markus A. Whiteland, and James Worrell. 2022. What's Decidable about Linear Loops?. *Proc. ACM Program. Lang.* 6, POPL, Article 65 (January 2022), 25 pages. https://doi.org/10.1145/3498727

1 INTRODUCTION

Loops are a fundamental staple of any programming language, and the study of loops plays a pivotal rôle in many subfields of computer science, including automated verification, abstract interpretation, program analysis, semantics, etc. The focus of the present paper is on the algorithmic analysis of simple (i.e., non-nested) linear (or affine) while loops, such as the following:

Authors' addresses: Toghrul Karimov, Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany; Engel Lefaucheux, Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany; Joël Ouaknine, Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany; David Purser, Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany; Anton Varonka, Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany and Saarbrücken Graduate School of Computer Science, Saarland Informatics Campus, Saarbrücken, Germany; Markus A. Whiteland, Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany; James Worrell, Department of Computer Science, University of Oxford, Oxford, UK.

2475-1421/2022/1-ART65
https://doi.org/10.1145/3498727

```
t := 1;
u := -1;
v := 2;
w := 0;
while (true) do
   t := 3t+2u-5w;
   u := u+3w;
   v := 4u+3v+w;
   w := t+u+2v;
```

In this running example, we make use of four numerical variables, t, u, v, w, initialised respectively to 1, -1, 2, and 0.[1] Note that all assignments in the body of the loop are linear functions of the variables; we therefore speak of a 4-*dimensional linear loop*.

The behaviour of the loop can be analysed (or 'specified') using logical formulas over a finite set of predicates on the variables. Several temporal-logic formalisms have been designed for this purpose; in this paper, we focus on Monadic Second-Order Logic (MSO)[2] over semialgebraic predicates,[3] which affords a high degree of expressiveness.[4]

In our running example, we could (for instance) define the following predicates:

$$P_1(t,u,v,w) : t + u + v - w = 0 \wedge (t^3 = u^2 \vee w \geq 3t^2 + u)$$
$$P_2(t,u,v,w) : t + u + 2v - 2w = 0 \wedge t^3 + v^2 + v > w$$
$$P_3(t,u,v,w) : t^4 - u^2 = 3 \wedge 2v^2 = w \wedge t^2 - 2u^3 = 4v$$

One can then express properties concerning the behaviour of the loop, such as the following:

$$\mathbf{G}(P_1 \Rightarrow \mathbf{F}\neg P_2) \wedge \mathbf{F}(P_3 \vee \neg P_1)\,.$$

For simplicity, the above formula is written in the language of LTL; in words, it asserts that whenever P_1 holds, then P_2 must eventually subsequently fail, and moreover that eventually either P_3 will hold or P_1 will fail (at some point in the execution of the loop).

The reader will probably agree that whether or not the above specification holds in our running example is not immediately obvious to determine (even, arguably, in principle). The main contribution of this paper is to exhibit an algorithm which can automatically decide the truth or falsity of such assertions. Informally speaking, we show decidability of the MSO model-checking problem for linear loops in which every predicate in the specification is a semialgebraic set that is *either* contained in some three-dimensional subspace, *or* has intrinsic dimension at most 1. We make these notions more precise shortly; for now, we simply remark that both P_1 and P_2 (viewed as semialgebraic subsets of $\mathbb{R}^4$) are each contained within three-dimensional subspaces, and P_3 has intrinsic dimension 1 (i.e., is 'string-like', or a curve, as a subset of $\mathbb{R}^4$). We also show that relaxing any of these constraints runs up against formidable mathematical obstacles (longstanding open

[1] We are purposely not specifying the exact types of the variables; mathematically speaking, in order for our analysis to be as general as possible, we shall assume that all variables range over the real numbers. In doing so, we are therefore deliberately not taking account of machine-implementation phenomena such as overflows or finite-precision rounding.

[2] Monadic Second-Order Logic is a highly expressive specification formalism that subsumes the vast majority of temporal logics employed in the field of automated verification, such as Linear Temporal Logic (LTL). See Section 2.2 for precise definitions.

[3] Semialgebraic predicates are Boolean combinations of polynomial equalities and inequalities.

[4] The use of predicates in the automated verification of software goes back some 25 years to the seminal work of Graf and Saïdi [Graf and Saïdi 1997]; however, most existing approaches and tools tend to limit themselves to Boolean combinations of linear (or affine) predicates (see, e.g., [Jhala et al. 2018]).

problems in number theory); our model-checking decidability result therefore appears to lie at the very frontier of what is achievable, barring major breakthroughs in mathematics.

Let us briefly comment on the scope of our setting. First, our focus is on *linear* loops, i.e., loops whose bodies contain exclusively linear assignments. However, affine assignments can also be handled, by including an extra (constant) variable and 'linearising' the loop; the net effect is to increase the dimension of the loop by 1. Second, program specifications occasionally make reference to the loop counter (i.e., the number of times the body of the loop has been executed); likewise, this can be accommodated within our framework simply by adding an extra variable, meant precisely to represent the loop counter, and increasing it by 1 on every iteration. Third, for simplicity we have dispensed with the guard condition in our running example; however any non-trivial guard can always be taken account of by including it as an additional predicate, so that specifications can refer to the point when the guard becomes false.[5]

We shall therefore focus on guard-free linear loops in the remainder of the paper. Such objects are in fact in one-to-one correspondence with *discrete-time linear dynamical systems*, and our subsequent exposition is therefore couched in the language of dynamical systems.

Dynamical systems are a fundamental modelling paradigm in many branches of science, and have been the subject of extensive research for many decades. A *discrete-time linear dynamical system* is given by a square $d \times d$ matrix M with rational entries, together with a starting point $x \in \mathbb{Q}^d$. The *orbit* of (M, x) is the infinite trajectory $\boldsymbol{O} = \langle x, Mx, M^2x, \dots \rangle$. The starting point x and matrix M corresponding to the loop in our running example are as follows:

$$x = \begin{pmatrix} 1 \\ -1 \\ 2 \\ 0 \end{pmatrix} \qquad M = \begin{pmatrix} 3 & 2 & 0 & -5 \\ 0 & 1 & 0 & 3 \\ 0 & 4 & 3 & 13 \\ 3 & 11 & 6 & 24 \end{pmatrix}.$$

Within the field of computer science, one of the earliest achievements concerning the analysis of linear dynamical systems is a celebrated result by Kannan and Lipton from the 1980s, the (polynomial-time) decidability of the *Orbit Problem* [Kannan and Lipton 1980, 1986]: given such a system (M, x), together with a point target $y \in \mathbb{Q}^d$, does the orbit of the system ever hit y?

Kannan and Lipton's paper answered an open problem of Harrison from the 1960s on reachability for linear sequential machines [Harrison 1969]. However, a secondary motivation was to propose an approach to attack the well-known Skolem Problem, which had itself been famously open since the 1930s (and remains unsolved to this day); phrased in the language of linear dynamical systems, the Skolem Problem asks whether it is decidable, given (M, x) as above, together with a $(d-1)$-dimensional subspace H of $\mathbb{R}^d$, to determine if the orbit of (M, x) ever hits H. This problem is known to be decidable in dimensions $d \leq 4$, and is otherwise open—for a more detailed discussion on the topic, we refer the reader to [Ouaknine and Worrell 2015]. Kannan and Lipton suggested that, in ambient space $\mathbb{R}^d$ of arbitrary dimension, the problem of hitting a low-dimensional subspace might be decidable. Indeed, this was eventually substantiated by Chonev *et al.* for linear subspaces of dimension at most 3 [Chonev et al. 2013, 2016].

Subsequent research focussed on the decidability of hitting targets of increasing complexity, such as half-spaces [Halava et al. 2006; Laohakosol and Tangsupphathawat 2009; Ouaknine and Worrell 2014a,b,c], polytopes [Almagor et al. 2017; Chonev et al. 2015; Tarasov and Vyalyi 2011], and semialgebraic sets [Almagor et al. 2019, 2021b]. Since discrete-time linear dynamical systems can equivalently be viewed as simple deterministic while loops with affine assignments, many of

[5] Note that we do not claim to be able to represent if-then-else conditionals in the loop body, which would immediately lead to undecidability, by reduction (for example) from the Halting Problem for two-counter machines.

the questions considered above also have immediate bearing on corresponding halting problems for such loops.

In recent years, motivated in part by verification problems for stochastic systems and linear loops, researchers have begun investigating more sophisticated specification formalisms than mere reachability: for example, the paper [Agrawal et al. 2015] studies approximate LTL model checking of Markov chains (which themselves can be viewed as particular kinds of linear dynamical systems), whereas [Karimov et al. 2020] focuses on LTL model checking of low-dimensional linear dynamical systems with semialgebraic predicates. In [Almagor et al. 2021a], the authors investigate the model-checking problem for diagonalisable linear dynamical systems in arbitrary dimension against prefix-independent MSO properties; both are significant restrictions—in particular, reachability queries are *not* prefix-independent and therefore do not fall within the scope of the problems considered in [Almagor et al. 2021a].

1.1 Main Contributions

In the present paper, we consider full MSO model checking of discrete-time linear dynamical systems of arbitrary dimension (that is, any number of program variables), only placing restrictions on the dimension of our semialgebraic predicates. More precisely, given a linear dynamical system (M, x) in ambient dimension d, together with a finite collection of (not necessarily disjoint) semialgebraic sets $\pi = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k\}$ (with each $\mathcal{T}_i \subseteq \mathbb{R}^d$), we associate with the orbit $\mathcal{O}$ of this dynamical system an infinite *characteristic word* $w(\mathcal{O}, \pi)$ over the 2^k-letter alphabet 2^π: writing $w(\mathcal{O}, \pi)[n]$ to denote the nth letter of this word, we require that

$$\mathcal{T}_i \in w(\mathcal{O}, \pi)[n] \text{ if and only if } M^n x \in \mathcal{T}_i\,.$$

In other words, $w(\mathcal{O}, \pi)$ keeps track at each discrete time step of the sub-collection of semialgebraic sets that the orbit is currently visiting.

In order to define our specification formalism, we formally associate to each subset $S \subseteq \pi$ of semialgebraic sets a unary predicate P_S, and require that $P_S(n)$ hold if and only if $w(\mathcal{O}, \pi)[n] = S$.

Our main result is as follows: provided that each of the semialgebraic sets in π *either* has intrinsic dimension at most 1, *or* is contained within some three-dimensional subspace of $\mathbb{R}^d$, the attendant MSO model-checking problem for discrete-time linear dynamical is decidable.[6]

Note that since we have a single starting point, the orbit consists of a single trajectory. The problem we are solving is sometimes referred to in the literature as "path checking", although typical applications in runtime verification and online monitoring involve finite traces, *e.g.*, [Leucker and Schallhart 2009]. Path checking ultimately periodic infinite traces is considered in [Markey and Schnoebelen 2003], but the traces arising from linear dynamical systems need not be ultimately periodic (see [Agrawal et al. 2015]).

To the best of our knowledge, our model-checking result goes substantially farther than existing work in the literature; indeed, even mere *reachability* (let alone MSO model checking!) of semialgebraic sets contained in 3D subspaces (or complements thereof, since MSO is closed under negation) was not previously known to be decidable. Moreover, further progress on the decidability of reachability appears to run into stringent difficulties. For instance, going beyond semialgebraic sets contained in 3D subspaces seems highly problematic. Indeed, the decidability of reachability

[6]The result is precisely stated in Theorem 3.1. The intrinsic dimension of a semialgebraic set is formally defined via cell decomposition; intuitively, one-dimensional semialgebraic sets can be viewed as 'strings' or 'curves'.

of a 4D polytope in four-dimensional ambient space presents formidable mathematical difficulties, as shown in [Chonev et al. 2015].[7] Furthermore, whether 3D polytopes (and hence, arbitrary semialgebraic sets contained within a four-dimensional subspace) can be reached is hard for the longstanding open case of the Skolem Problem at order 5. The same applies for the reachability of semialgebraic sets of intrinsic dimension 2 [Baier et al. 2021].

Our decidability result rests primarily on the following:

Theorem 1.1. *[Semenov 1984, Theorem 1] For any effectively almost periodic word w, the MSO theory of $(\mathbb{N}, <)$ expanded with unary predicates that define w is decidable.*

Our approach therefore consists in establishing that the characteristic word associated with the orbit of a linear dynamical system, given our dimensional constraints on the semialgebraic predicates, is always 'effectively almost periodic' (the precise definition is provided later in the paper). In order to achieve this, we make extensive use of spectral techniques, as well tools recently developed in [Baier et al. 2021] for handling parametric linear dynamical systems.

1.2 Relevance to PL and Related Work

It is perhaps worth emphasising once more the manner in which our work relates to the field of Programming Languages, broadly construed. We have already pointed out that linear loops are in one-to-one correspondence with discrete-time linear dynamical systems, and therefore model-checking and verification questions pertaining to the latter immediately carry over to the former.

Linear loops have been extensively studied and play an important role in the foundations of program analysis and software verification. In particular, methods to prove termination—or, equivalently, reachability—via a variety of techniques, such as spectral analysis or the synthesis of ranking functions, have been developed in, e.g., [Ben-Amram et al. 2019; Ben-Amram and Genaim 2013, 2014, 2017; Bradley et al. 2005; Braverman 2006; Chen et al. 2015; Colón and Sipma 2001; Hosseini et al. 2019; Podelski and Rybalchenko 2004a,b; Tiwari 2004]. Several of these approaches have been implemented in software verification tools, such as Microsoft's Terminator [Cook et al. 2006a,b].

Other research directions include the development of acceleration techniques for linear loops, together with the automated synthesis of closed forms, as in [Boigelot 2003; Jeannet et al. 2014; Kincaid et al. 2019], as well as the automated generation of invariants, e.g., [Almagor et al. 2018; Colón et al. 2003; Cousot 2005; Cousot and Halbwachs 1978; Fijalkow et al. 2019; Gupta et al. 2008; Kincaid et al. 2018; Lefaucheux et al. 2021; Ouaknine and Worrell 2015; Rodríguez-Carbonell and Kapur 2004, 2007].

Model-checking algorithms, such as the one presented in this paper, both complement and enhance the toolkit of techniques enabling the automated analysis of program loops and, by extension, more complex pieces of code.

2 PRELIMINARIES

2.1 Words

Given a finite alphabet Σ, let Σ^* be the set of finite words over Σ and $\Sigma^{\mathbb{N}}$ be the set of infinite words over Σ. Given a word w, let $w[i]$ be the i'th character in w, indexed from 0, i.e. $w = w[0]w[1]w[2]\cdots$. We say that a non-empty word $u = u_0 \cdots u_j$ *occurs at position* k in w if $w[k]w[k+1]\cdots w[k+j] = u$.

[7] In a nutshell, such a decidability result would entail substantial breakthroughs in the field of Diophantine approximation.

2.2 Monadic Second-Order Logic (MSO)

Temporal logics specify the required behaviour of the system as it evolves with time. Typically these are specified as ω-regular properties (regular languages over infinite words). Such properties can be captured in Monadic Second-Order Logic (MSO) over the structure $(\mathbb{N}, <)$ (representing the ordered set of positions of an infinite word) and a finite collection of predicates $P_1, \ldots, P_k : \mathbb{N} \to \{\texttt{true}, \texttt{false}\}$. These predicates are used as indicators as to whether the location orbit is within a target at time i. Second-order quantification is permitted, but is restricted to quantification over sets of positions. The grammar of a monadic second-order specification is as follows[8]:

$\psi := P(i)$	(where $P(i)$ is a predicate on position i of the word)
$\psi := \exists i \in \mathbb{N} : \psi \mid \forall i \in \mathbb{N} : \psi$	(first-order quantification)
$\psi := \exists X \subseteq \mathbb{N} : \psi \mid \forall X \subseteq \mathbb{N} : \psi$	(subset quantification/monadic second-order quantification)
$\psi := i \in X \mid i \notin X$	(subset membership testing)
$\psi := i < j \mid i = j$	(index comparison)
$\psi := \neg\psi \mid \psi \vee \psi \mid \psi \wedge \psi \mid \psi \Rightarrow \psi$	(standard logical operations)
$\psi := i = 0 \mid i = 1 \mid i = 2 \mid \ldots$	(fixed values)

We are interested in model checking MSO on infinite words. Here the natural numbers represent positions in the word, and the predicates indicate sets of positions in a word, for example, $P_S(i)$ could be defined to indicate whether $w(\mathcal{O}, \pi)[i] = S$.

Example 2.1. Examples of MSO formulas for model checking LDS:

- Reachability of target $\mathcal{T}_i$: $\exists n : P_{\mathcal{T}_i}(n)$.
- Eventually trapped inside $\mathcal{T}_i$: $\exists n \forall m : m > n \implies P_{\mathcal{T}_i}(m)$.
- In target $\mathcal{T}_i$ at every odd position (O is the set of odd natural numbers in the following): $\exists O \subseteq \mathbb{N} : 1 \in O \wedge \forall x \in O, \exists y, z : (y \notin O \wedge z \in O \wedge x < y < z \wedge \nexists t : x < t < y \vee y < t < z) \wedge \forall x : x \in O \implies P_{\mathcal{T}_i}(x)$.
- Whenever $\mathcal{T}_i$ is visited $\mathcal{T}_j$ is visited some point later: $\forall n : P_{\mathcal{T}_i}(n) \implies \exists m > n : P_{\mathcal{T}_j}(m)$.
- Any linear temporal logic (LTL) formula over predicates $P_{\mathcal{T}_1}, \ldots, P_{\mathcal{T}_m}$.

Decomposing the predicates. We will use the following simple observation on several occasions. Consider two sets of semi-algebraic predicates $\pi = \{\mathcal{T}_1, \ldots, \mathcal{T}_k\}$ and $\pi' = \{\mathcal{T}_1', \ldots, \mathcal{T}_{k'}'\}$. Suppose that for all $i \in \{1, \ldots, k\}$, each set $\mathcal{T}_i$ can be written as a Boolean combination of $\mathcal{T}_1', \ldots, \mathcal{T}_{k'}'$. Then the problem of model checking MSO formulas on the characteristic word $w(\mathcal{O}, \pi)$ can be reduced to that of model checking MSO formulas on $w(\mathcal{O}, \pi')$.

2.3 Algebraic Numbers and Eigenvalues

Our linear dynamical systems are defined using rational matrices, however our techniques rely on the analysis of eigenvalues and the Jordan normal form of a matrix. Given a matrix M, the roots of the characteristic equation $\det(M - \lambda I) = 0$ are the eigenvalues. In general the eigenvalues of a rational matrix, and thus the entries of the Jordan normal form are not necessarily rational, but algebraic. The set of algebraic numbers $\overline{\mathbb{Q}}$ comprises those complex numbers that are roots of univariate polynomials with rational coefficients. In particular, rational numbers are algebraic numbers. For every $\alpha \in \overline{\mathbb{Q}}$ there exists a unique monic univariate polynomial p_α with rational coefficients of minimum degree for which $p_\alpha(\alpha) = 0$. We call p_α the *minimal polynomial* of α. An algebraic number α is represented as a tuple (p, a, ε), where p is its minimal polynomial, $a = a_1 + a_2 i$,

[8]Some expressions can be written as combination of the others, but are included to make the expressivity clear.

with $a_1, a_2 \in \mathbb{Q}$, is an approximation of α, and $\varepsilon \in \mathbb{Q}$ is sufficiently small such that α is the unique root of p within distance ε of a (such ε can be computed by the root-separation bound, due to Mignotte [Mignotte 1982]). This is referred to as the *standard* or *canonical representation* of an algebraic number. Given canonical representations of algebraic numbers α and β, one can compute canonical representations of $\alpha + \beta$, $\alpha\beta$, and α/β, all in polynomial time (see e.g., [Ouaknine et al. 2017, Section 2.4]).

Let $\mathbb{T}$ be the set of points on the unit torus. That is the set of complex numbers with modulus one. Such a number α is a root of unity if $\alpha^n = 1$ for some n.

2.4 Linear Recurrence Sequences

Linear dynamical systems and linear recurrent sequences (LRS) are strongly related. An order-k linear recurrence sequence (LRS) $\langle u_n \rangle_{n\in\mathbb{N}}$ is computed by $u_n = a_1 u_{n-1} + \cdots + a_k u_{n-k}$, for fixed $a_1, \ldots, a_k \in \mathbb{R}$ and initial values $u_1, \ldots, u_k \in \mathbb{R}$. In fact, the sequence of values of each coordinate of a linear dynamical system (or program variables of a linear loop), can be described by a linear recurrence sequence. We will make use of this strong interdependence and first recall some key properties of LRS.

Its characteristic polynomial is $p(x) = x^k - a_1 x^{k-1} - \ldots - a_{k-1}x - a_k$, and further the roots of p are called the *characteristic roots* of the LRS. An LRS with characteristic roots $\lambda_1, \ldots, \lambda_t$ can be expressed in closed form $u_n = p_1(n)\lambda_1^n + \cdots + p_t(n)\lambda_t^n$, for polynomials $p_1, \ldots, p_t$ with degrees depending on the multiplicities of the roots and coefficients on the initial values $u_1, \ldots, u_k$. We refer the reader to [Kauers and Paule 2011, Chapter 4] for further reading on this topic.

2.5 Almost Periodic Words and Arc-Hitting Models

Given a target $\mathcal{T} \subseteq \mathbb{R}^d$ we define the hitting set as the times the orbit hits (is contained within) $\mathcal{T}$ that is $\mathcal{Z}(\mathcal{T}) = \{n \mid M^n x \in \mathcal{T}\}$. For the targets of the form considered in this paper we aim to characterise such sets, for which, we introduce the following definitions.

Definition 2.2. An infinite word $w \in \Sigma^{\mathbb{N}}$ over the alphabet Σ is called *periodic* if there exists $p > 0$ such that $w[n] = w[n + p]$ for all $n \geq 0$. An infinite word w is *eventually periodic* if there exist N and $p > 0$ such that $w[n] = w[n + p]$ for all $n \geq N$.

Eventually periodic words can be represented by a finite directed graph, where each node has exactly one successor and each edge is labelled by a character. However, in some cases we must generalise this notion as follows:

Definition 2.3. An infinite word $w \in \Sigma^{\mathbb{N}}$ is *almost periodic* if for any finite word $u \in \Sigma^*$, there exists $B_u \in \mathbb{N}$ such that either u does not occur in w after position B_u, or the gap between any two consecutive occurrences of u is at most B_u.

The word w is *effectively almost periodic* if B_u can be computed for any given word u.

Our goal is to show that the characteristic word is effectively almost periodic. To do this we will represent sets using so-called arc-hitting models. For some intuition, consider a circle with an arc covering some part of the perimeter. An arrow points from the center of the circle to its edge, and is allowed to move only by rotating by a fixed angle. This system represents a set $\mathcal{Z} \subseteq \mathbb{N}$ defined so that the integer n is in $\mathcal{Z}$ if and only if the nth rotation of the arrow points into the arc. Formally we represent this model using open subsets of the unit circle in the complex plain (or torus) $\mathbb{T} := \{z \in \mathbb{C} \colon |z| = 1\}$:

Definition 2.4. An arc on the torus $\mathbb{T}$ is a connected subset of $\mathbb{T}$ and is defined by three points: it is the circular arc which connects its two endpoints and passes through the third point. An open

arc is an arc whose endpoints are removed. A set $\mathcal{Z} \subseteq \mathbb{N}$ is represented by an *arc-hitting model* if there exist an algebraic number λ with $|\lambda| = 1$ but which is not a root of unity[9], a finite union I of open arcs on $\mathbb{T}$, $N \in \mathbb{N}$, and a finite set $F \subseteq \{0, \ldots, N-1\}$ such that $n \in \mathcal{Z}$ if and only if either $\lambda^n \in I$ and $n \geq N$, or $n \in F$. The *angle* of an arc-hitting model is the argument of λ, $\arg(\lambda)$.

The indicator word $w \in \{0,1\}^{\mathbb{N}}$ ($w[n] = 1$ if and only if $n \in \mathcal{Z}$) of a set $\mathcal{Z}$ represented by an arc-hitting model is known to be almost periodic [Muchnik et al. 2003, Theorem 15]. The arc-hitting model doesn't capture all sets with almost periodic indicator words, but is sufficient for our purposes; for example, they do not capture sets with eventually periodic indicator words because such sets would require roots of unity (which we exclude). Arc-hitting models can also represent finite and cofinite sets (X is cofinite if there exists $N \in \mathbb{N}$ such that $X \cup \{0, \ldots, N\} = \mathbb{N}$)—here $I = \emptyset$ or $\mathbb{T}$ respectively, with F taking care of the finite part.

2.6 The Point Target Case

Let us first consider the case where a target $\mathcal{T} \subseteq \mathbb{R}^d$ is in fact a single point, that is, $\mathcal{T} = \{t\}$ for $t \in \mathbb{R}^d$. In this case we observe that $\mathcal{Z}(\mathcal{T})$ is either finite or eventually periodic. In fact, if a point t is repeated then the whole orbit is eventually periodic, since the dynamics of the system between the two occurrences will repeat indefinitely, and we can revert to model checking an eventually periodic word. Two applications of the Kannan-Lipton orbit problem can detect this case: first ask if (M, x) reaches $\mathcal{T}$ (if not, $\mathcal{Z}(\mathcal{T})$ is empty), and if the first hitting time is n, then ask if $(M, M^{n+1}x)$ hits $\mathcal{T}$, if so the system is eventually periodic and otherwise $\mathcal{Z}(\mathcal{T}) = \{n\}$.

Our analysis will show that for certain 1D semialgebraic targets, only a finite number of the points from the target can be reached—in which case these targets reduce to a finite union of points and can be handled by the preceding analysis.

3 DEGENERACY

For the sake of further analysis, we account for degeneracy. A LDS is *degenerate* if there exist two distinct eigenvalues λ_i, λ_j of matrix M such that their quotient λ_i/λ_j is a root of unity.

As captured in [Everest et al. 2003, Section 1.1.9], the study of degenerate LDS can be reduced to that of finitely many related non-degenerate systems. The idea is to decompose the orbit of a degenerate system specified by a matrix M into $L \in \mathbb{N}$ many non-degenerate suborbits, each specified by matrix M^L. We take this approach, and so will need to put the suborbits back together again—we do this in Section 6.

We take a short detour towards the Jordan normal form to discuss the eigenvalues of matrix M and its powers. It is well-known that there exists a nonsingular matrix S such that $M = S^{-1}JS$, where $J = \operatorname{diag}(J_1, \ldots, J_t)$ is a block diagonal matrix. Each block of J has the following form:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}, \tag{1}$$

where $\lambda_1, \ldots, \lambda_t$ are (not necessarily distinct) eigenvalues of M. A block J_i is referred to as a *Jordan block* of λ_i. Matrix J is the *Jordan normal form*, or simply *Jordan form*, of M. Since J is block diagonal, so is any power: $J^m = \operatorname{diag}(J_1^m, \ldots, J_t^m)$. As a matter of fact, the diagonal entries of a

[9] A number α is a root of unity if there exists n such that $\alpha^n = 1$. In particular, if λ is not a root of unity, the argument of λ is not a rational multiple of π.

Jordan block power J_i^m are all λ_i^m. Hence, $\lambda_1^m, \ldots, \lambda_t^m$ are all eigenvalues of J^m. Since $M^m = S^{-1}J^mS$, we conclude that the eigenvalues of M^m are the same as eigenvalues of J^m.

In order to determine the matrix-specific value of L mentioned previously, we first define the set of all eigenvalue quotients of M which are roots of unity:

$$\Omega = \{\lambda_i/\lambda_j : \lambda_1, \ldots, \lambda_t \text{ eigenvalues of } M \text{ and } \lambda_i/\lambda_j \text{ is a root of unity}\}.$$

Given a root of unity ω, let order(ω) be the smallest positive integer such that $\omega^{\text{order}(\omega)} = 1$, and let $L = \text{lcm}\{\text{order}(\omega) : \omega \in \Omega\}$. We note that every power of M that is a multiple of L has no degenerate eigenvalues. Let λ_i, λ_j be two eigenvalues of M such that λ_i^L and λ_j^L are distinct eigenvalues of M^L. The ratio λ_i^L/λ_j^L is not a root of unity: indeed, if $\lambda_i^L/\lambda_j^L = (\lambda_i/\lambda_j)^L$ is a root of unity, then so is λ_i/λ_j. In that case, order(λ_i/λ_j) divides L by definition and this implies a contradiction $\lambda_i^L = \lambda_j^L$. We recall that all eigenvalues of M^L are L-th powers of M's eigenvalues and conclude that M^L is non-degenerate. For the same reason, since M^L is non-degenerate, so is M^{kL} for any fixed positive integer k. Moreover, the matrix M^{kL} is not only non-degenerate; each of its non-real eigenvalue is not a real multiple of a root of unity. To see that, let λ be an eigenvalue of M such that $\lambda^{kL} = A\omega$, where $A \in \mathbb{R}$ and ω is a root of unity. Since λ^{kL} is non-real, its conjugate $\overline{\lambda^{kL}}$ is an eigenvalue of M^{kL} as well. Then, however, their ratio $\lambda^{kL}/\overline{\lambda^{kL}} = A\omega/A\omega^{-1} = \omega^2$ would be a root of unity, contradicting non-degeneracy. Finally, by setting $L := \text{lcm}\{L, 2\}$, we can assume that all multiples of roots of unity among eigenvalues of M^L are *positive* reals.

We can then consider L subsequences $((M^L)^nM^rx)_{n=0}^{\infty}$, for each $r \in \{0, \ldots, L-1\}$. Notice that each instance uses the same matrix M^L, but the starting points differ.

Our approach is to describe, for every target $\mathcal{T}$ and every $r \in \{0, \ldots, L-1\}$ the set $\mathcal{Z}_r(\mathcal{T})$ of n such that $(M^L)^nM^rx \in \mathcal{T}$. We will show that $\mathcal{Z}_r(\mathcal{T})$ can be represented with arc-hitting models (including possibly because $\mathcal{Z}_r(\mathcal{T})$ is finite or cofinite).

From the sets $\mathcal{Z}_r(\mathcal{T})$, we will be able to reconstruct the characteristic word $w(O, \pi)$ in an effectively almost periodic way (Theorem 6.1), and hence apply Theorem 1.1 to prove our main theorem:

Theorem 3.1. *Consider a linear dynamical system (M, x), a collection π of semialgebraic targets each of which either has intrinsic dimension at most one or linear dimension at most three, and an MSO formula ψ. Let $w(O, \pi)$ be the characteristic word. Then it is decidable whether the characteristic word $w(O, \pi)$ satisfies ψ.*

Here the *intrinsic dimension* of a semialgebraic set is the usual dimension of a semialgebraic set (see [Bochnak et al. 1998] and Section 5.1). The *linear dimension* of a semialgebraic set S is the dimension of the subspace spanned by the points in S. A subspace of dimension k can be defined by k independent vectors $v_1, \ldots, v_k$ as $\{v_1a_1 + \cdots + v_ka_k \mid a_1, \ldots, a_k \in \mathbb{R}\}$.

Henceforth, when constructing $\mathcal{Z}_r(\mathcal{T})$ in Section 5 and Section 4 we consider only the non-degenerate matrix M^L.

4 SEMIALGEBRAIC TARGETS CONTAINED IN 3D SUBSPACES

In this section, we show that given an update matrix M, a starting point x and a semialgebraic target set $\mathcal{T}$ contained inside a 3D subspace[10], it is possible to partition the orbit $\langle x, Mx, M^2x, \ldots\rangle$ into L subsequences such that the set of time steps at which each of the L suborbits enters $\mathcal{T}$ can be described using an arc-hitting model. We then use this result in Theorem 6.1, which shows effective

[10] $\mathcal{T}$ is contained in a 3D subspace if there exist linearly independent v_1, v_2, v_3 such that $\mathcal{T}$ is contained in the span of v_1, v_2, v_3.

almost-periodicity of infinite words obtained by interleaving words that can be described using arc-hitting models.

Theorem 4.1. *Let $M \in \mathbb{Q}^{d\times d}$ and $x \in \mathbb{Q}^d$. For every semialgebraic target set $\mathcal{T} \subseteq \mathbb{R}^d$ contained inside a 3D subspace V of $\mathbb{R}^d$, there exists $L > 0$ such that for $0 \le r < L$, $\mathcal{Z}_r(\mathcal{T}) = \{n \in \mathbb{N} : M^{nL+r}x \in \mathcal{T}\} = \{n \in \mathbb{N} : (M^L)^n(M^r x) \in \mathcal{T}\}$ can be described by an arc-hitting model.*

Theorem 4.1 shows that the sequence $(M^n x)_{n\in\mathbb{N}}$ can be written as an interleaving of L sequences that can be described using arc-hitting models. Observe that it suffices to prove Theorem 4.1 for non-degenerate matrices M. To see this, suppose that the theorem holds for non-degenerate systems and let M be an arbitrary matrix and D be such that M^D is non-degenerate. Define sequences $x^0, \dots, x^{D-1}, (x_i^k) = M^{iD+k}x$. Since M^D is non-degenerate, each x^i can be written as an interleaving of L_i sequences described by arc-hitting models. Observing that if a sequence can be described using L_i arc-hitting models, then it can be described using L_i' arc-hitting models for every multiple L_i' of L_i, we can take the least common multiple of $L_0, \dots, L_{D-1}$ and assume that $L_i = L_j = L'$ for every i, j. Next, define $L = L'D$ and consider the subsequences $y^0, \dots, y^{L-1}$,where $y_i^k = M^{iL+k}x$. We have to construct an arc-hitting model for each y^i. Wlog consider y^0. Observe that y^0 is a subsequence of x^0: in fact, $y_i^0 = x_{iL'}^0$ for $i \ge 0$. By assumption, x^0 can be described (more precisely, the set $\{n \in \mathbb{N} : x_n^0 \in \mathcal{T}\}$ can be described) using an arc-hitting model with parameters N, λ, I. Then y^0 can be described by the "L'-times accelerated" arc-hitting model with parameters NL', $\lambda^{L'}$ and I.

To prove Theorem 4.1, we begin by investigating $\mathcal{Z}(V)$ where V is a subspace.

Theorem 4.2. *Let V be a subspace of $\mathbb{R}^d$ and (M, x) a linear dynamical system. Then $\mathcal{Z}(V)$ is semilinear, that is, of the form $F \cup \bigcup_{i=1}^{s}(r_i + N\mathbb{N})$ for finite F and arithmetic progressions $r_i + N\mathbb{N}$ for $1 \le i \le s$.*

Proof. If $V = \mathbb{R}^d$, then $\mathcal{Z}(V) = \mathbb{N}$, which is semilinear. Otherwise, V can be written as an intersection $V_1 \cap \ldots \cap V_m$ of m hyperplanes of dimension $d-1$. By the Skolem-Mahler-Lech Theorem [Hansel 1985], $\mathcal{Z}(V_i)$ is semilinear for each i, and intersection of semilinear sets remains semilinear. □

The Skolem-Mahler-Lech Theorem, however, is not constructive in the sense that it does not give us a way to construct the semilinear sets $\mathcal{Z}(V_1), \dots, \mathcal{Z}(V_m)$. In fact, the famously open Skolem Problem boils down to deciding whether the finite set F is empty. We will show (in Theorem 4.5) that it is possible to write down the set $\mathcal{Z}(V)$ if $\dim(V) \le 3$. To do this, we will need the following lemmata. First, Theorem 4.3 which combines the results of [Chonev et al. 2016, Lemmata G.1, G.3-G.4]:

Lemma 4.3. *Let $(u_i)_{i\in\mathbb{N}}$ be a non-zero non-degenerate LRS of order at most 4. If*

(1) $u_n = A\lambda_1^n + \overline{A}\,\overline{\lambda_1^n} + B\lambda_2^n + \overline{B}\,\overline{\lambda_2^n}$, *or*
(2) $u_n = (A + Bn)\lambda_1^n + (\overline{A} + \overline{B}n)\overline{\lambda_1^n}$

for $A, B, \lambda_1, \lambda_2 \in \overline{\mathbb{Q}}$, then there exists a computable bound N on the set $\{n \in \mathbb{N} : u_n = 0\}$.

Understanding zeros of LRS (in particular, computing effective bounds on zeros of LRS) is interesting to us due to the connection to reachability problems for linear dynamical systems. For example, given a hyperplane H, there exists an LRS u such that $M^n x \in H$ if and only if $u_n = 0$. One way to obtain such bounds for LRS of the form $u_n = v_n + w_n$ is to compute a lower bound on $|v_n|$. For example, if $|v_n| > \frac{1}{p(n)}$ for some polynomial $p(n)$, and $|w_n| < (1-\epsilon)^n$ for some $\epsilon > 0$, then u_n cannot be 0 for large n. This is exactly how the following lemma will be used later.

Lemma 4.4. *Let $\lambda, c \in \overline{\mathbb{Q}}$ with $|\lambda| = 1$, λ not a root of unity and $B \in \mathbb{R} \cap \overline{\mathbb{Q}}$. There exist computable values $D, N > 0$ such that for all $n > N$, $|c\lambda^n + \overline{c}\,\overline{\lambda^n} + B| > \frac{1}{n^D}$.*

Proof. It suffices to only consider the case where $|c| = 1$, as one can show that D, N satisfy the statement of the lemma for λ, c, B if and only if D', N satisfies the statement of the lemma for $\lambda, \frac{c}{|c|}, B$ where $D' > 0$ is sufficiently large with respect to $|c|$.

Since $c\lambda^n + \overline{c}\,\overline{\lambda^n}$ takes values in $[-2, 2]$, if $|B| > 2$, then $|c\lambda^n + \overline{c}\,\overline{\lambda^n} + B|$ is bounded below by a positive constant and the conclusion follows immediately. Henceforth we assume that $B \in [-2, 2]$.

Define $f(z) = |z + \overline{z} + B| = |2\,\mathrm{Re}(z) + B|$. Then $|c\lambda^n + \overline{c}\,\overline{\lambda^n} + B| = f(c\lambda^n)$.

Given the restriction $B \in [-2, 2]$, $f(z)$ will have exactly two conjugate zeroes in the unit circle $\mathbb{T}$, which we denote with w and $\overline{w}$. Using [Ouaknine and Worrell 2014a, Corollary 8] we can compute D, N such that for all $n > N$, $|c\lambda^n - w|, |c\lambda^n - \overline{w}| > \frac{1}{n^{D/2}}$. We show that this implies that for all $n > N$, $|f(c\lambda^n)| > \frac{1}{n^D}$. We begin by writing

$$|f(c\lambda^n)| = |f(c\lambda^n) - f(w)| = |c\lambda^n + \overline{c}\,\overline{\lambda^n} - w - \overline{w}| = 2|\,\mathrm{Re}(c\lambda^n) - \mathrm{Re}(w)|.$$

Recall that for all $n > N$, $|c\lambda^n - w|, |c\lambda^n - \overline{w}| > \frac{1}{n^{D/2}}$. By considering the geometry of the unit circle,

$$|z_1 - z_2| > \frac{1}{n^{D/2}} \implies |\,\mathrm{Re}(z_1) - \mathrm{Re}(z_2)| > |\,\mathrm{Re}(1) - \mathrm{Re}(e^{i\frac{1}{n^{D/2}}})| = 1 - \cos\frac{1}{n^{D/2}}.$$

We hence obtain that for all $n > N$,

$$2|\,\mathrm{Re}(c\lambda^n) - \mathrm{Re}(w)| > 2|(1 - \cos\frac{1}{n^{D/2}})| > \frac{1}{n^D}. \qquad \square$$

We are now ready to analyse $\mathcal{Z}(V)$ where V is a linear subspace of dimension at most three.

Theorem 4.5. *Let V be a linear subspace of $\mathbb{R}^d$ with $\dim V \leq 3$, and (M, x) a non-degenerate linear dynamical system with $M \in \mathbb{Q}^{d\times d}$ and $x \in \mathbb{Q}^d$. Either*

- *$\mathcal{Z}(V) = \mathbb{N}$, or*
- *$\mathcal{Z}(V)$ is finite with an effectively computable upper bound N on the elements of $\mathcal{Z}(V)$.*

Proof. Observe that whether $\mathcal{Z}(V) = \mathbb{N}$ can be determined by simply checking whether the first four elements x, Mx, M^2x, M^3x of the orbit are in V. To see this, suppose $x, Mx, M^2x, M^3x \in V$. Then there must exist $k \leq 3$ such that $M^kx = c_0x + \ldots + c_{k-1}M^{k-1}x$ for some $c_0, \ldots, c_{k-1} \in \mathbb{R}$. Multiplying both sides of the linear dependence equation by powers of M, we conclude that for all $i \geq 3$ (in fact, for all $i \geq k$), M^ix can be written as a linear combination of the k preceding elements $M^{i-k}x, \ldots, M^{i-1}x$. Since the subspace V is closed under taking linear combinations, using induction we can conclude that $M^ix \in V$ for all $i \geq 0$.

We next show how to compute a bound on $\mathcal{Z}(V)$ in case $\mathcal{Z}(V) \neq \mathbb{N}$. In this proof we exploit the *real Jordan normal* form to streamline the arguments. Let $\lambda_1, \ldots, \lambda_k$ be the real eigenvalues of M, while $\lambda_{k+1}, \ldots, \lambda_s$ are non-real (and closed under conjugation). The real Jordan normal form is a block diagonal matrix as in the definition given in Section 3. A *real Jordan block* R_i corresponding to $\lambda_i, i = 1, \ldots, k$ is defined as in (1), whereas the Jordan blocks $J_1, \ldots, J_\ell$ of complex eigenvalues are redefined. Every *complex Jordan block* of the form J_j corresponds to a pair of conjugate complex eigenvalues $\lambda_j = a_j + b_ji$, $\overline{\lambda_j} = a_j - b_ji$ and has the following form:

$$J_j = \begin{bmatrix} \Lambda_j & I & & \\ & \Lambda_j & \ddots & \\ & & \ddots & I \\ & & & \Lambda_j \end{bmatrix}, \quad \Lambda_j = \begin{bmatrix} a_j & -b_j \\ b_j & a_j \end{bmatrix}$$

for $1 \leq j \leq l$, where $a_j, b_j \in \mathbb{R} \cap \overline{\mathbb{Q}}$. The benefit of working with the real Jordan form $J = \mathrm{diag}(R_1, \ldots, R_k, J_1, \ldots, J_\ell)$ is that all matrices involved are over $\mathbb{R} \cap \overline{\mathbb{Q}}$ rather than $\overline{\mathbb{Q}}$. Let m be the multiplicity of J_j, that is, $J_j \in \mathbb{R}^{2m \times 2m}$. Powers of J_j have the following form:

$$J_j^n = \begin{bmatrix} \Lambda_j^n & \binom{n}{1}\Lambda_j^{n-1} & \binom{n}{2}\Lambda_j^{n-2} & \cdots & \binom{n}{m-1}\Lambda_j^{n-m+1} \\ & \Lambda_j^n & \binom{n}{1}\Lambda_j^{n-1} & \cdots & \binom{n}{m-2}\Lambda_j^{n-m+2} \\ & & \ddots & \ddots & \vdots \\ & & & \Lambda_j^n & \binom{n}{1}\Lambda_j^{n-1} \\ & & & & \Lambda_j^n \end{bmatrix}, \quad \Lambda_j^n = \begin{bmatrix} \mathrm{Re}(\lambda^j) & -\mathrm{Im}(\lambda^j) \\ \mathrm{Im}(\lambda^j) & \mathrm{Re}(\lambda^j) \end{bmatrix}. \tag{2}$$

Observe that J_j^m consists of blocks that are a polynomial (in n) multiples of powers of Λ_j.

We can assume that M is in real Jordan form since any problem instance can be translated into one with the update matrix in this form by observing that for any S, J such that $M = S^{-1}JS$, $M^n x \in V$ if and only if $J^n(Sx) \in S(V)$ where $S(V)$ is the image of V under the coordinate transform S with $\dim S(V) = \dim V$. It will also be convenient to assume that for $1 \leq i \leq k$, the entry of x that corresponds to the bottom row of R_i (i.e. the entry of x that, when computing in Mx, will be multiplied by the diagonal entry in the bottom row of R_i) is non-zero and for $1 \leq j \leq l$, the two coordinates of x that correspond to the bottom two rows of J_j are not both zero. Any given instance with M, x and V can be transformed into this form by removing equations corresponding to certain coordinates that are always zero and modifying V accordingly. For example, suppose $M = \begin{bmatrix} \Lambda & I & 0 \\ 0 & \Lambda & I \\ 0 & 0 & \Lambda \end{bmatrix} \in \mathbb{R}^{6\times 6}$ and $x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & 0 & 0 \end{bmatrix}^\top \in \mathbb{R}^6$. Then $M^n x \in V$ if and only if $\begin{bmatrix} \Lambda & I \\ 0 & \Lambda \end{bmatrix}^n \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}^\top \in W$, where $W = \{v \in \mathbb{R}^4 : (v, 0, 0) \in V\}$. In particular, $\dim(W) \leq \dim V$.

Next we show how to compute the bound N on $\mathcal{Z}(V)$ by a case analysis on the structure of J. Case I considers problem instances whose update matrix has at least four (i.e. at least two conjugate pairs of) non-real eigenvalues. The idea of the proof is to first show that the "global" condition $M^n x \in V$ is satisfied only if a certain local condition is satisfied by the nth powers of two blocks of M that, between them, have four different non-real eigenvalues. We then show that this sufficient local condition already enforces that $\mathcal{Z}(V)$ is finite. Cases II and III partition the problem instances with at most one conjugate pair of non-real eigenvalues based on whether $\mathrm{diag}(J_1, \ldots, J_l)$ is simple or not. These two cases could also be handled by using the facts that the Skolem-Mahler-Lech Theorem is effective when M has at most one pair of non-real eigenvalues (see [Tijdeman et al. 1984, Theorem 1]), and that V can be written as an intersection of hyperplanes.

Case I. Suppose M has four distinct non-real eigenvalues, i.e. there exist blocks J_i and J_j with eigenvalues $\lambda_i, \overline{\lambda_i}, \lambda_j, \overline{\lambda_j}$ and $\Lambda_i \neq \Lambda_j$. We denote the entries of x corresponding to the bottom two rows of J_i and J_j by x_1, x_2 and x_3, x_4 respectively, with $(x_1, x_2) \neq \mathbf{0}$ and $(x_3, x_4) \neq \mathbf{0}$ by the assumption discussed above. Let W be the projection of V onto the coordinates that correspond to $x_1, \ldots, x_4$ and $H = \{x \in \mathbb{R}^4 : c^\top x = 0\}$, $c^\top = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix} \neq \mathbf{0}$ a hyperplane that contains W. Such H must exist as $\dim W \leq 3$ (this is where the assumption that $\dim V \leq 3$ is crucial). We have

$$M^n x \in V \implies \begin{bmatrix} a_i & -b_i & 0 & 0 \\ b_i & a_i & 0 & 0 \\ 0 & 0 & a_j & -b_j \\ 0 & 0 & b_j & a_j \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in W \implies [c_1\, c_2\, c_3\, c_4] \begin{bmatrix} a_i & -b_i & 0 & 0 \\ b_i & a_i & 0 & 0 \\ 0 & 0 & a_j & -b_j \\ 0 & 0 & b_j & a_j \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0.$$

By analysing the powers of Λ_i and Λ_j, the rightmost condition can be written as

$$A\lambda_i^n + \overline{A}\,\overline{\lambda_i^n} + B\lambda_j^n + \overline{B}\,\overline{\lambda_j^n} = 0$$

where $\lambda_i = a_i + b_i i$, $\lambda_j = a_j + b_j i$, $A = c_1(x_1 - x_2 i) + c_2(x_1 i + x_2)$ and $B = c_3(x_3 - x_4 i) + c_4(x_3 i + x_4)$. Since all the variables are real-valued, $c \neq \mathbf{0}$, and at least one of (x_1, x_2) and (x_3, x_4) is not equal to $\mathbf{0}$, A and B both cannot be zero. Hence $u_n = A\lambda_i^n + \overline{A}\,\overline{\lambda_i^n} + B\lambda_j^n + \overline{B}\,\overline{\lambda_j^n}$ is a real-valued, non-degenerate linear recurrence sequence, and by Theorem 4.3, it has finitely many zeros with a computable bound N such that $u_n \neq 0$ for $n > N$. Going back to (M, x) and V, we can conclude that $\mathcal{Z}(V)$ is also bounded by N.

Case II. There exists a block J_j with multiplicity at least 2 (i.e. with at least 4 rows). Similarly to the preceding case, considering only the four coordinates corresponding to the bottom four rows of J_j, define x_1, x_2, x_3, x_4 and project V onto the 4 relevant coordinates to obtain W. Let $H = \{x \in \mathbb{R}^4 : c^\top x = 0\}$, $c \neq \mathbf{0}$ be a hyperplane that contains W. We have

$$M^n x \in V \implies \begin{bmatrix} a_i & -b_i & 1 & 0 \\ b_i & a_i & 0 & 1 \\ 0 & 0 & a_i & -b_i \\ 0 & 0 & b_i & a_i \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in W \implies [c_1\ c_2\ c_3\ c_4] \begin{bmatrix} a_i & -b_i & 1 & 0 \\ b_i & a_i & 0 & 1 \\ 0 & 0 & a_i & -b_i \\ 0 & 0 & b_i & a_i \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0.$$

The last equation can be written as

$$C(n)\lambda_i^n + \overline{C(n)}\,\overline{\lambda_i^n} = 0 \tag{3}$$

where $\lambda_i = a_i + b_i i$ and

$$C(n) = c_1(x_1 + x_2 i) + c_2(x_1 i + x_2) + c_3(x_3 + x_4 i) + c_4(x_3 i + x_4) + \frac{c_1(x_3 + i x_4) + c_2(-x_3 i + x_4)}{\lambda_i} n.$$

Recalling that x_3, x_4 are not both zero and that all variables are real-valued, we observe that $c_1(x_3 + i x_4) + c_2(-x_3 i + x_4) \neq 0$ and conclude that $C(n)$ is not identically zero. We can therefore write Equation 3 as $(A + Bn)\lambda_i^n + (\overline{A} + \overline{B}n)\overline{\lambda_i^n} = 0$, $A, B \in \overline{\mathbb{Q}}$, which by Theorem 4.3, has a computable upper bound on the solutions in n.

Case III. $J_i = \Lambda = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ for every i. In this case, M has at most one pair of non-real eigenvalues, $\lambda = a + bi$ and $\overline{\lambda} = a - bi$, and all blocks with non-real eigenvalues have multiplicity 1. We show that for every hyperplane $H = \{x \in \mathbb{R}^d : c^\top x = 0\}$, $c \neq \mathbf{0}$, $\mathcal{Z}(H)$ is finite and can be effectively bounded. Hence either $V = \mathbb{R}^d$, in which case $\mathcal{Z}(V) = \mathbb{N}$, or V is contained in a hyperplane H and $\mathcal{Z}(V) \subset \mathcal{Z}(H)$ can be effectively bounded.

By considering powers of M (e.g., see Equation 2) and observing that $\mathrm{Re}(\lambda^n) = \frac{1}{2}(\lambda^n + \overline{\lambda^n})$ and $\mathrm{Im}(\lambda^n) = \frac{1}{2}(\lambda^n - \overline{\lambda^n})$,

$$c^\top M^n x = C\lambda^n + \overline{C}\,\overline{\lambda^n} + \sum_{i=1}^{k} p_i(n)\rho_i^n$$

where $C \in \overline{\mathbb{Q}}$, $\rho_1, \ldots, \rho_k$ are the real eigenvalues of M with $\rho_1 > \cdots > \rho_k > 0$ (recall from Section 3 that we can assume that all real eigenvalues are positive and distinct), and $p_i(n)$ are polynomials with non-zero real algebraic coefficients for $1 \leq i \leq k$.

Let $\sigma = \max\{|\lambda|, \rho_1, \ldots, \rho_k\}$ be the spectral radius of M, $\gamma = \frac{\lambda}{\sigma}$ and $\delta_i = \frac{\rho_i}{\sigma}$ for $1 \leq i \leq k$. Observe that either $|\gamma| = 1$, or $\delta_1 = 1$ and $|\gamma|, \delta_2, \ldots, \delta_k < 1$. Define

$$v_n = \frac{c^\top M^n x}{\sigma^n} = C\gamma^n + \overline{C}\,\overline{\gamma^n} + \sum_{i=1}^{k} p_i(n)\delta_i^n.$$

We have that $M^n x \in H \iff c^\top M^n x = 0 \iff v_n = 0$. Hence it suffices to show that the set of all zeros of v_n can be effectively bounded.

- If $|\gamma| < 1$ (in which case, $\delta_1 = 1$), then $|p_1(n)\delta_1^n| = \Omega(1)$ (i.e. asymptotically bounded from below by a positive constant), whereas $|C\gamma^n + \overline{C}\,\overline{\gamma^n}|$ and $|\sum_{i=2}^k p_i(n)\delta_i^n|$ decrease exponentially to 0. Hence we can compute N such that for all $n > N$, $|p_1(n)\delta_1^n| > |C\gamma^n + \overline{C}\,\overline{\gamma^n} + \sum_{i=2}^k p_i(n)\delta_i^n|$ and hence $v_n \neq 0$.
- Similarly, if $\delta_1 < 1$ (in which case, $|\gamma| = 1$), then $|\sum_{i=1}^k p_i(n)\delta_i^n|$ decreases exponentially to 0 with n, whereas by Theorem 4.4 we can compute a constant D and a bound N such that for all $n > N$, $|C\gamma^n + \overline{C}\,\overline{\gamma^n}| > \frac{1}{n^D}$. Hence we can compute a bound N' such that for all $n > N'$, $|C\gamma^n + \overline{C}\,\overline{\gamma^n}| > \frac{1}{n^D} > \left|\sum_{i=1}^k p_i(n)\right|$ and hence $v_n \neq 0$.
- If $\delta_1 = |\gamma| = 1$ and $p_1(n)$ is not constant, then $|p_1(n)\delta_1^n|$ goes to infinity with n, whereas $|C\gamma^n + \overline{C}\,\overline{\gamma^n}|$ is bounded by a constant and $|\sum_{i=2}^k p_i(n)\delta_i^n|$ decreases exponentially. Hence we can compute N such that for all $n > N$, $|p_1(n)\delta_1^n| > |C\gamma^n + \overline{C}\,\overline{\gamma^n} + \sum_{i=2}^k p_i(n)\delta_i^n|$ and hence $v_n \neq 0$.
- If $\delta_1 = |\gamma| = 1$ and $p_1(n) = B$ is constant, then by Theorem 4.4 we can compute a constant D and a bound N such that for all $n > N$, $|C\gamma^n + \overline{C}\,\overline{\gamma^n} + B| > \frac{1}{n^D}$. Hence we deduce that for all $n > N$, $|C\gamma^n + \overline{C}\,\overline{\gamma^n} + p_1(n)\delta_1^n| = |C\gamma^n + \overline{C}\,\overline{\gamma^n} + B| > \frac{1}{n^D}$. Since $|\sum_{i=2}^k p_i(n)\delta_i^n|$ decreases exponentially to 0, we can then compute a bound $N' > N$ such that for all $n > N'$, $|C\gamma^n + \overline{C}\,\overline{\gamma^n} + p_1(n)\delta_1^n| > \frac{1}{n^D} > \left|\sum_{i=2}^k p_i(n)\right|$ and hence $v_n \neq 0$. □

The preceding theorem describes $\mathcal{Z}(V)$ for a 3D subspace V. Recall that we are interested in understanding $\mathcal{Z}(\mathcal{T})$ for $\mathcal{T}$ contained inside such a subspace. To this end, we will combine Theorem 4.5 with the following result about three-dimensional dynamical systems from [Karimov et al. 2020].

Lemma 4.6. *Let $\mathcal{T} \subseteq \mathbb{R}^3$ be a semialgebraic set and (M, x) a non-degenerate dynamical system with $M \in (\overline{\mathbb{Q}} \cap \mathbb{R})^{3\times 3}$ and $x \in (\overline{\mathbb{Q}} \cap \mathbb{R})^3$. $\mathcal{Z}(\mathcal{T})$ can be described by an arc-hitting model.*

Proof. Let $p : \mathbb{R}^3 \to \mathbb{R}$ be a polynomial in three variables. The approach is to show that

- the set $\{n : p(M^n x) = 0\}$ is either finite or equal to $\mathbb{N}$, and
- the sets $\{n : p(M^n x) > 0\}$ and $\{n : p(M^n x) > 0\}$ each can be described by an arc-hitting model with λ that only depends on M.

This suffices because the semialgebraic set $\mathcal{T}$ can be defined as a Boolean combination of sets $\mathcal{T}_1, \ldots, \mathcal{T}_k$, where for $1 \le i \le k$, $\mathcal{T}_i = \{(x_1, x_2, x_3) : p_i(x_1, x_2, x_3) \sim_i 0\}$ and $\sim_i \in \{<, =, >\}$. Hence $\mathcal{Z}(\mathcal{T})$ is a Boolean combination of $\mathcal{Z}(\mathcal{T}_1), \ldots, \mathcal{Z}(\mathcal{T}_k)$, which are either finite or cofinite or can be described using an arc-hitting model with the same λ. It remains to observe that finite and cofinite sets can be represented by arc-hitting models (with any parameter λ that is not a root of unity), and that taking a Boolean combination of arc-hitting models with the same parameter λ yields a single arc-hitting model (with parameter λ).

First consider the case where M has only real eigenvalues $\rho_1, \rho_2, \rho_3 \in \mathbb{R}$ and let $J = SMS^{-1}$ be the Jordan form of M. Let $x_1(n), x_2(n), x_3(n)$ denote the three coordinates of $M^n x$. By analysing $M^n x = S^{-1} J^n S x$ we can observe that for $1 \le i \le 3$, $x_i(n)$ is of the form

$$x_i(n) = \sum_{j=1}^{3} q_j(n)\rho_j^n$$

where for $1 \leq j \leq 3$, q_j is a polynomial with real algebraic coefficients. Since $p(M^n x)$ is obtained from $x_1(n), x_2(n), x_3(n)$ through multiplication and addition, $p(M^n x)$ will be of the form

$$\sum_{i,j,k<K} q_{i,j,k}(n)\rho_1^{in}\rho_2^{jn}\rho_3^{kn} = \sum_{i,j,k<K} q_{i,j,k}(n)\rho_{i,j,k}^{n}$$

for some $K > 0$, polynomials $q_{i,j,k}$ with algebraic coefficients and real algebraic $\rho_{i.j.k}$. It remains to observe that expressions of this type are either identically zero, ultimately positive or ultimately negative.

Now suppose M has a complex eigenvalue λ (which, by the assumption of non-degeneracy, cannot be a root of unity). [Karimov et al. 2020, Section 4] shows, again by writing M in Jordan form and considering powers of M, that there exists a computable N such that for all $n > N$,

$$\operatorname{sign}(p(M^n x)) = \operatorname{sign}\left(\sum_{m=0}^{K} \beta_m \gamma^{nm} + \overline{\beta_m \gamma^{nm}} + r(n)\right)$$

where sign maps $\mathbb{R}$ to $\{-, 0, +\}$, $\gamma = \frac{\lambda}{|\lambda|}$, $\beta_m \in \overline{\mathbb{Q}}$ for all m and $r(n)$ decreases exponentially to 0 as $n \to \infty$. From this we can deduce that either $\operatorname{sign}(p(M^n x))$ is always 0 (in case $\beta_m = 0$ for all m), or [Karimov et al. 2020, Theorem 5] that there exist computable open subsets $I_>$ and $I_<$ of $\mathbb{T}$ (that are finite unions of open intervals) such that for all $n > N$, $p(M^n x) \neq 0$ and $p(M^n x) \sim 0$ if and only if $\gamma^n \in I_\sim$. This gives us arc-hitting models with parameters N, γ and one of $I_>, I_<$. □

Proof of Theorem 4.1. Recall that it suffices to prove Theorem 4.1 for non-degenerate M. Let $M \in \mathbb{Q}^{d\times d}$, $x \in \mathbb{Q}^d$, V be a 3D subspace of $\mathbb{R}^d$ and $\mathcal{T} \subseteq V$ a semialgebraic target. Consider $\mathcal{Z}(V)$. By Theorem 4.5, there are two possibilities. If $\mathcal{Z}(V)$ is finite with an effectively computable upper bound, then so is $\mathcal{Z}(\mathcal{T})$ and we can describe $\mathcal{Z}(\mathcal{T})$ using a single arc-hitting model (i.e. $L = 1$).

Now suppose $\mathcal{Z}(V) = \mathbb{N}$, i.e. the orbit of (M, x) always remains inside V. In this case, (M, x) is essentially a three-dimensional dynamical system. More formally, let $D \leq 3$ be the maximal number of independent vectors in $\{x, Mx, M^2x\}$. We can then define a D-dimensional "projected" dynamical system (M_p, x_p) and a semialgebraic set $\mathcal{T}_p \subseteq \mathbb{R}^D$ such that $M^n x \in \mathcal{T} \iff M_p^n x_p \in \mathcal{T}_p$. For example, if $D = 3$, then using $\{x, Mx, M^2x\}$ as a basis for $\mathbb{R}^3$ we can write $x_p = \begin{bmatrix}1 & 0 & 0\end{bmatrix}$, $M_p = \begin{bmatrix}Mx & | & M^2x & | & M^3x\end{bmatrix}$ and $\mathcal{T}_p = \{(a, b, c) : ax + bMx + cM^2x \in \mathcal{T}\}$. But now observe that we can characterize the set of all n such that $M_p^n x_p \in \mathcal{T}_p$ using Theorem 4.6. Let L be such that M_p^L is non-degenerate. Then, by Theorem 4.6, we have that $\mathcal{Z}_r(\mathcal{T}_p)$ (and hence $\mathcal{Z}_r(\mathcal{T})$) can be described by an arc-hitting model for $0 \leq r < L$. □

5 1D SEMIALGEBRAIC TARGETS

In this section we consider semialgebraic target sets that have (intrinsic) dimension 1 but need not stay within a 3D subspace of $\mathbb{R}^d$. Such sets are essentially finite unions of curves, and in Section 5.1 we show that a 1D semialgebraic target $\mathcal{T}$ can be represented as union $\mathcal{T} = \bigcup_{i=1}^{\ell}\{v_i(s) : s \in \mathbb{R}\}$ of sets parametrized by an algebraic function of a single variable.

Let M be an update matrix, $L > 0$ such that M^L is non-degenerate, x a starting point and $\mathcal{T}$ a semialgebraic target of dimension 1 (which we will formally define shortly). Similarly to the main theorem of Section 4, we will show that for $0 \leq r < L$, $\mathcal{Z}_r(\mathcal{T}) = \{n \in \mathbb{N} : M^{nL+r}x \in \mathcal{T}\}$ can be described by an arc-hitting model. We extend the work of [Baier et al. 2021], which (implicitly) showed how to determine whether $\mathcal{Z}(\mathcal{T})$ is empty or not, by fully characterising the set using arc-hitting models.

Theorem 5.1. *Given a non-degenerate $(M, \tilde{x})$ and a semialgebraic target $\mathcal{T}$ of dimension 1, the set $\mathcal{Z}(\mathcal{T})$ can be effectively represented with arc-hitting models.*

Assuming non-degeneracy matrices, we further observe that we can assume the problem to be given in Jordan form, as defined in Section 3. Suppose $M = S^{-1}JS$ with J in Jordan form and that $\mathcal{T} = \{v(s) : s \in \mathbb{R}\}$. Then $M^n x \in \mathcal{T} \iff J^n(Sx) = Sv(s)$ for some s. Let $\tilde{x} = Sx$ and $\tilde{v}(s) = Sv(s)$. Here $\tilde{v}(s)$ is a linear transformation on $v(s)$ and therefore is also an algebraic function. Notice that the entries of $J, \tilde{x}$ and $\tilde{v}$ may be non-real due to J being the complex Jordan form. The remainder of this section discusses 1D semialgebraic sets and proves Theorem 5.1 with the translation to JNF in mind.

5.1 Expressing Semialgebraic Sets Parametrically

Dimension of a semialgebraic set is defined using Cell Decomposition (see, e.g., [Bochnak et al. 1998, Chapter 2]). In particular, a semialgebraic set of dimension one (in our case, the target $\mathcal{T}$) is a union of cells $C_1, \ldots, C_k$ of dimension one in $\mathbb{R}^d$, which can be defined inductively as follows.

- Cells of dimension one in $\mathbb{R}$ are either points in $\overline{\mathbb{Q}} \cap \mathbb{R}$, or open intervals with endpoints in $\overline{\mathbb{Q}} \cup \{-\infty, \infty\}$;
- A cell C of dimension one in $\mathbb{R}^{k+1}$ with $k > 0$ can be written as $C = \{(x, g(x)) : x \in D\}$, where D is a cell of dimension one in $\mathbb{R}^k$ and $g(x)$ is the unique value of y satisfying the system $p_1(x, y) = 0, q_1(x, y) > 0, \ldots, q_m(x, y) > 0$ where $p_1, q_1, \ldots, q_m$ are polynomials in $k + 1$ variables with integer coefficients. In other words, $C \subseteq \mathbb{R}^{k+1}$ is the image of the semialgebraic function g over the cell $D \subseteq \mathbb{R}^k$ of dimension one.

Consider a 1D semialgebraic target $\mathcal{T} \subseteq \mathbb{R}^d$. We show that $\mathcal{T}$ can be written as a finite union of sets parametrized by an algebraic function over $\mathbb{R}$, i.e. sets of the form $\{f(s) : s \in \mathbb{R}\}$ where $f : \mathbb{R} \mapsto \mathbb{R}^d$ is an algebraic function. We do this by using induction to show that each cell of dimension one can be written as a union of sets parametrized by an algebraic function over $\mathbb{R}$. For the base case, observe that a point $p \in \overline{\mathbb{Q}} \cap \mathbb{R}$ can be characterized using the algebraic function $f(s) = p$, the interval $(0, 1]$ as $\{\frac{1}{1+s^2} : s \in \mathbb{R}\}$ and the interval $(0, \infty)$ as $\{\frac{1}{s^2} : s \in \mathbb{R}\}$. We can characterize all other types of intervals using these parametrizations. For example, $(a, b] = \{a + \frac{b-a}{1+s^2}\}$, $[b, a) = \{a - \frac{a-b}{1+s^2} : s \in \mathbb{R}\}$ and an open interval (a, b) can be written as $(a, b) = (a, \frac{a+b}{2}] \cup [\frac{a+b}{2}, b)$.

Next, let C be a cell of dimension 1 in $\mathbb{R}^{k+1}$. Recall that

$$C = \{(x, y) : x \in D, p_1(x, y) = 0, q_1(x, y) > 0, \ldots, q_m(x, y) > 0\}$$

where D is a cell of dimension 1 in $\mathbb{R}^k$. By the induction hypothesis, D must be a union of sets $D_1, \ldots, D_\ell$ parametrized by $f_1(s), \ldots, f_\ell(s)$, respectively. Hence $C = \bigcup_{i=1}^{\ell} C_i$, where for $1 \le i \le \ell$,

$$C_i = \{(f_i(s), y) : s \in \mathbb{R},\ p_1(f_i(s), y) = 0,\ q_1(f_i(s), y) > 0,\ \ldots,\ q_m(f_i(s), y) > 0\}$$

is the component of C that is obtained from D_i. We need to show that each C_i can be parametrized by an algebraic function. Viewing $p_1(f_i(s), y)$ and $q_j(f_i(s), y)$, $1 \le j \le m$, as polynomials in y with coefficients that are algebraic functions of s, we can factorise to obtain the system

$$\begin{cases} p_1(f_i(s), y) = (y - h_1^0(s)) \cdot \ldots \cdot (y - h_{\kappa(0)}^0(f_1(s))) = 0 \\ q_1(f_i(s), y) = (y - h_1^1(s)) \cdot \ldots \cdot (y - h_{\kappa(1)}^1(f_1(s))) > 0 \\ \ldots \\ q_m(f_i(s), y) = (y - h_1^m(s)) \cdot \ldots \cdot (y - h_{\kappa(m)}^m(f_1(s))) > 0 \end{cases}$$

where h_r^i is an algebraic function for every $0 \le i \le m$ and $1 \le r \le \kappa(i)$. Next we will show how to compute $\kappa(0)$ subsets $I_1, \ldots, I_{\kappa(0)}$ of $\mathbb{R}$ that have the following properties.

- $\bigcup_{j=1}^{\kappa(0)} I_j = \mathbb{R}$;
- Each I_j is a finite union of intervals;

- For $1 \le j \le \kappa(0)$, for all $s \in I_j$ the value of y corresponding to $f_i(s)$ is equal to $h_j^0(s)$ (that is, $(f_i(s), h_j^0(s)) \in C_i$), which is the jth root of $p_1(f_i(s))$.

That is, I_j is the set of all values of s for which $y = h_j^0(s)$. This will allow us to write

$$C_i = \bigcup_{j=1}^{\kappa(0)} \{(f_i(s), h_j^0(s)) : s \in I_j\}.$$

Here each I_j is a finite union of intervals and hence can be parametrized using algebraic functions with domain $\mathbb{R}$. Since composition of two algebraic functions remains algebraic, we can characterize each component of C_i that comes from a single subinterval of I_j using algebraic functions with domain $\mathbb{R}$. Hence we can write C_i as a union of sets with the desired parametrization.

To construct I_j for each $1 \le j \le \kappa(0)$, we proceed as follows. Since $I_j = \{s : (f_i(s), h_j^0(s)) \in C_i\}$, it can be defined by the formula

$$\varphi(s) \coloneqq p_1(f_i(s), h_j^0(s)) = 0 \wedge q_1(f_i(s), h_j^0(s)) > 0 \wedge \cdots \wedge q_m(f_i(s), h_j^0(s)) > 0.$$

Hence I_j is semialgebraic. Since semialgebraic sets have finitely many connected components, I_j must be a finite union of interval subsets of $\mathbb{R}$.

5.2 Non-diagonalisable M

Having shown that we can consider the target $\mathcal{T}$ as an image of an algebraic function, towards Theorem 5.1 we now show that $\mathcal{Z}(\mathcal{T})$ can be represented using an arc-hitting model. We begin with the case that M is non-diagonalisable and show that in this case $\mathcal{Z}(\mathcal{T})$ is in fact either finite or co-finite. We consider the case where M is diagonalisable in Section 5.3 .

Let $J = (J_1, \ldots, J_t)$ be the Jordan form of the non-degenerate matrix, where J_i is a Jordan block of dimension d_i corresponding to an eigenvalue λ_i. We index $\tilde{x}$ by $\tilde{x}_{i,1}, \ldots, \tilde{x}_{i,d_i}$ for the coordinates corresponding to J_i (where $\tilde{x}_{i,1}$ corresponds to the bottom row), similarly for $\tilde{v}(s)$. For example:

$$J = \begin{bmatrix} J_1 & & \\ & J_2 & \\ & & J_3 \end{bmatrix} \quad \tilde{x} = \begin{bmatrix} \tilde{x}_{1,1} \\ \tilde{x}_{2,d_2} \\ \vdots \\ \tilde{x}_{2,1} \\ \tilde{x}_{3,d_3} \\ \vdots \\ \tilde{x}_{3,1} \end{bmatrix} \quad \tilde{v}(s) = \begin{bmatrix} \tilde{v}_{1,1}(s) \\ \tilde{v}_{2,d_2}(s) \\ \vdots \\ \tilde{v}_{2,1}(s) \\ \tilde{v}_{3,d_3}(s) \\ \vdots \\ \tilde{v}_{3,1}(s) \end{bmatrix}$$

Lemma 5.2. *Given non-degenerate J, if there exists a non-diagonal Jordan block J_i with eigenvalue λ_i, then $\mathcal{Z}(\mathcal{T})$ is effectively finite or cofinite.*

Proof. Recall from the non-degeneracy assumption of Section 3 that either λ_i is not a root of unity or $\lambda_i = 1$.

If λ_i is not a root of unity, we can apply [Baier et al. 2021, Lemma 18] which shows there is an effective bound on n for which $\lambda_i^n \tilde{x}_{i,1} = \tilde{v}_{i,1}(s)$ and $\lambda_i^n \tilde{x}_{i,2} + n\lambda_i^{n-1} \tilde{x}_{i,1} = \tilde{v}_{i,2}(s)$ can both hold. This entails that $\mathcal{Z}(\mathcal{T})$ is finite.

Now let us turn to the case when $\lambda_i = 1$, we will still conclude $\mathcal{Z}(\mathcal{T})$ is finite or cofinite. We consider polynomial equations in variable n, formed by $J_i^n \tilde{x}_i = \tilde{v}_i(s)$, for all Jordan blocks with $\lambda_i = 1$. For J_i, this leads to constraints on (n, s) in the following form:

$$\tilde{x}_{i,1} = \tilde{v}_{i,1}(s), \quad n\tilde{x}_{i,1} + \tilde{x}_{i,2} = \tilde{v}_{i,2}(s), \quad \ldots, \quad \sum_{j=1}^{d_i} \binom{n}{d_i - j} \tilde{x}_{i,j} = \tilde{v}_{i,d_i}(s).$$

Equations which do not depend on n, for example $\tilde{x}_{i,1} = \tilde{v}_{i,1}(s)$ must either hold for all s, or there are finitely many choices of s because there are only finitely many roots of an algebraic function (reducing the problem to single point targets).

For the remaining equations, there is an equation of the form $n = (\tilde{v}_{i,2}(s) - \tilde{x}_s)/\tilde{x}_{i,1}$, which can be used to replace n in all other equations. Again we test whether the constraint system is satisfied for all s or only finitely many s (in which case we again reduce to single point targets).

If the constraints hold for all s, we have the equation $n = (\tilde{v}_{i,2}(s) - \tilde{x}_s)/\tilde{x}_{i,1}$. If the range $\{(\tilde{v}_{i,2}(s) - \tilde{x}_s)/\tilde{x}_{i,1} \mid s \in R\}$ is bounded then we conclude that $\mathcal{Z}(\mathcal{T})$ is finite.

Otherwise, if the constraints do not hold for all s, we take an equation of the form $\lambda_j^n \tilde{x}_{j,1} = \tilde{v}_{j,1}(s)$ from some Jordan block with eigenvalue different from 1 (if it exists) and can again apply [Baier et al. 2021, Lemma 18] to bound n such that $\lambda_j^n \tilde{x}_{j,1} = \tilde{v}_{j,1}(s)$ and $n = (\tilde{v}_{i,2}(s) - \tilde{x}_s)/\tilde{x}_{i,1}$ both hold, concluding that $\mathcal{Z}(\mathcal{T})$ is finite. If no such equation exists (because all Jordan blocks have eigenvalue 1) then n is unbounded, and $\mathcal{Z}(\mathcal{T})$ is cofinite. □

5.3 Diagonalisable M

In the remainder, we complete the proof of Theorem 5.1 when the matrix is diagonalisable, and so we have constraints of the form $\lambda_i^n \tilde{x}_i = \tilde{v}(s)_i$, $i = 1, \ldots, t$. Henceforth, we rewrite this as $\lambda_i^n = \gamma_i(s)$, where $\gamma_i(s) = \tilde{v}(s)_i/\tilde{x}_i$. In order to do this, we must assume that $\tilde{x}_i \neq 0$. Observe that if $\tilde{x}_i = 0$ (or indeed $\lambda_i = 0$), then the constraint can be dropped: either all n satisfy the constraint if there exists $s : \tilde{v}(s)_i = 0$, or otherwise no n satisfy the constraint and $\mathcal{Z}(\mathcal{T})$ is empty.

Eigenvalues can either be real or complex. We recall, due to our non-degeneracy assumption, that no complex eigenvalue has rational argument (that is a rational multiple of 2π). That is there are no real multiples of roots of unity, except the positive reals.

Hence, if any eigenvalue λ_i is a root of unity then $\lambda_i = 1$, forming the constraint $1^n = 1 = \gamma_i(s)$. This holds either at finitely many s (reducing $\mathcal{T}$ to point targets), or $\gamma_i(s) = 1$ identically (for all s) in which case the constraint holds for all n trivially. Thus we may assume that no λ_i is a root of unity.

In the remainder of this section, we assume that no eigenvalue $\lambda_1, \ldots, \lambda_t$ is 0 or 1 by removing such equations as described above.

We split our case analysis depending on whether there exist two *multiplicatively independent* eigenvalues, that is, whether there exists i, j such that $\lambda_i^a \neq \lambda_j^b$ for all $a, b \in \mathbb{Z}$ not both zero.[11] Indeed, if there are two multiplicatively independent eigenvalues, then the following lemma of [Baier et al. 2021] entails that $\mathcal{Z}(\mathcal{T})$ is finite.

Lemma 5.3 ([Baier et al. 2021, Lemma 20]). *Suppose λ_1, λ_2 are constant, not roots of unity, and are multiplicatively independent. Assume further that γ_1, γ_2 are non-constant algebraic functions. Then the system $\lambda_1^n = \gamma_1(s)$, $\lambda_2^n = \gamma_2(s)$ has only finitely many solutions, and there is an effectively computable upper bound on such n.*

It remains that all pairs of eigenvalues are multiplicatively dependent. In particular, for each pair λ_i, λ_j, we have $\lambda_i^{a_2} = \lambda_j^{a_1}$ for some integers a_1, a_2 not both zero. In fact, since we assume that no eigenvalue of interest is a root of unity, we have that neither a_1 nor a_2 equals 0.

We observe that we cannot have both reals and complex numbers because we have eliminated the case where λ_i's may be (real multiples of) roots of unity. Suppose λ_i is complex and λ_j is real, but then $\lambda_i^{a_1} = \lambda_j^{a_2}$ implies $\lambda_i^{a_1}$ is real (thus λ_i is a (real multiple of a) root of unity).

[11] Given a collection $Y = \{\lambda_1, \ldots, \lambda_t\}$ of algebraic numbers consider the set $L = \{(a_1, \ldots, a_t) \in \mathbb{Z}^t : \lambda_1^{a_1} \cdots \lambda_t^{a_t} = 1\}$. It forms an abelian group under component-wise addition, and a deep result of Masser [Masser 1988] shows that a basis of L can be computed (in polynomial time, see, e.g., [Cai et al. 2000]). In particular, it is decidable whether any two of the eigenvalues λ_i are multiplicatively independent.

Secondly, observe that a non-real λ_i must be of modulus 1. Suppose we have complex λ_i, then since M is real we also have the complex conjugate $\lambda_j = \overline{\lambda_i}$. As λ_i and $\overline{\lambda_i}$ are multiplicatively dependent, we have $\lambda_i^{a_1} = \overline{\lambda_i}^{a_2}$ for $a_1, a_2 \neq 0$. Then $\lambda_i^{a_1+a_2} = \lambda_i^{a_2}\overline{\lambda_i}^{a_2} = (|\lambda_i|^2)^{a_2}$. Hence either $|\lambda| = 1$ or $a_1 = a_2$. However if $a_1 = a_2$ then $\lambda_i^{a_1} = \overline{\lambda_i}^{a_1}$ is real and so λ_i is a (real multiple of a) root of unity, which we have already excluded.

All real. First we suppose that every λ_i is real; as mentioned in Section 3 we may further assume that λ_i is non-negative. We will show that $\mathcal{Z}(\mathcal{T})$ is either finite or cofinite.

We have for any two eigenvalues $\lambda_i^{a_i} = \lambda_j^{a_j}$ and require $\lambda_i^n = \gamma_i(s)$ and $\lambda_j^n = \gamma_j(s)$. Since $\lambda_i = \lambda_j^{a_j/a_i}$, we have $\gamma_i(s) = \lambda_i^n = (\lambda_j^{a_j/a_i})^n = (\lambda_j^n)^{a_j/a_i} = \gamma_j(s)^{a_j/a_i}$. We either have $\gamma_i(s) = \gamma_j(s)^{a_j/a_i}$ holds identically, in which case we can drop one of the equations, or there are finitely many such s. In which case this reduces to the single point target problem. Hence we only need to worry about a single equation, let us assume this is $\lambda_i^n = \gamma(s)$.

Further since $\gamma(s)$ only crosses 0 finitely many times, some can partition $\mathcal{T}$ by splitting into regions of R where $\gamma(s)$ is of constant sign. We have $\mathcal{Z}(\mathcal{T}) = \emptyset$ whenever $\lambda_i > 0$ and $\gamma(s) \leq 0$, so we assume $\gamma(s) > 0$.

Hence we solve $\lambda^n = \gamma(s)$ for $\lambda, \gamma > 0$. Now, suppose $\lambda = 1$, either we have $\gamma = 1$, in which case $\mathcal{Z}(\mathcal{T}) = \mathbb{N}$. Or $\gamma(s) = 1$ for finitely many s, in which case we partition $\mathcal{T}$ into finitely many point targets.

The case where $\lambda > 1$ can be reduced to $\lambda < 1$ by considering $(1/\lambda)^n = 1/\gamma(s)$ if necessary. Thus assume $\lambda < 1$. Then if $\inf_{s\in R} \gamma(s) > 0$, we have $\lambda^n < \inf_{s\in R} \gamma(s)$ for some n and so $\mathcal{Z}(\mathcal{T})$ is finite. If $\inf_{s\in R} \gamma(s) = 0$ we reach $\mathcal{T}$ by the intermediate value theorem for every $n \geq m$ where $\lambda^m < \sup_{s\in R} \gamma(s)$. Hence $\mathcal{Z}(\mathcal{T})$ is cofinite.

Some non-real. Let us first work under the assumption there is a single constraint $\lambda^n = \gamma(s)$, where λ is a complex number of modulus one, but not a root of unity. Therefore λ^n takes on values in the unit circle. In case $\gamma(s)$ intersects the unit circle only at finitely many s, the problem reduces to the finitely many targets case. Otherwise the range of $\gamma(s)$, by continuity, is a finite union of arcs on the unit circle. Thus we can construct an arc-hitting model with λ and these arcs.

In general we have several constraints. Just like in the all real case, we want to reduce to a single constraint. However, one has to be careful when taking rational powers $\lambda_i^{a_i/a_j}$ or $\gamma_i(s)^{a_i/a_j}$, as these are multivalued functions over the complex numbers. In the analysis that follows we indeed show that the problem reduces to a single constraint one, but the arguments are quite technical due to the intricacies of complex exponentiation.

We express the pairwise multiplicative dependencies between the eigenvalues through λ_1. For the sake of readability, we set $\lambda = \lambda_1$ and $\gamma = \gamma_1$. Then, for each $j = 2, \ldots, t$, let a_j and b_j be non-zero integers such that $\lambda^{a_j} = \lambda_j^{b_j}$, $j = 2, \ldots, t$. Let $\ell = \operatorname{lcm}\{b_2, \ldots, b_t\}$. Fix μ to be one of the ℓ complex numbers in the set $\lambda^{1/\ell}$. Take then an algebraic function η satisfying $\eta^\ell = \gamma$ identically, with η continuous over D (e.g., a suitable root of the polynomial obtained by replacing y with y^ℓ in the minimal polynomial of γ).

The aim is to show that, for each $0 \leq r < \ell$, the set of solutions (n, s), with $n = r \mod \ell$, to the original system of equations is exactly the union of solutions to the equations $\mu^n = \omega\eta(s)$, where ω ranges over a suitable subset (depending on r) of the ℓth roots of unity.

To this end, notice that $\mu^{\ell a_j} = \lambda^{a_j} = \lambda_j^{b_j}$ so we have $\omega_j \mu^{\ell a_j/b_j} = \lambda_j$ for some b_jth root of unity ω_j. Furthermore, the number $c_j := \ell a_j/b_j$ is an integer. Similarly we have $\eta^{\ell a_j}(s) = \gamma^{a_j}(s) = \gamma_j^{b_j}(s)$

for all $s \in D$, so we have $\gamma_j(s) = \omega'_{s,j}\eta^{c_j}(s)$ for each s and some b_jth root of unity $\omega'_{s,j}$. In fact, $\omega'_{s,j}$ is constant in s since η and γ are continuous. Thus we may write $\gamma_j = \omega'_j\eta^{c_j}$.

We plug these relations into the constraints $\lambda_j^n = \gamma_j(s)$, $j = 1, \ldots, t$ to obtain the equations $\omega_j^n \mu^{c_j n} = \omega'_j \eta^{c_j}$. Notice that these equations are not "constant" in n in the sense that ω_j^n varies with n, but we can take arithmetic progressions with period ℓ to get "constant" equations. That is to say, for each r, $0 \le r < \ell$, and for each $n = r \mod \ell$, we have $\lambda_j^n = \omega_j^n \mu^{c_j n} = \omega_j^r \mu^{c_j n}$. The equation $\lambda_j^n = \gamma_j(s)$ is then equivalent to $\mu^{c_j n}\omega_j^r = \omega'_j \eta_j^{c_j}(s)$. Writing $\omega''_j = \omega'_j/\omega_j^r$ for each j (with r fixed), we obtain the following equivalent system of equations

$$\begin{cases} \mu^{\ell n} & = \eta^\ell(s) \\ \mu^{c_j n} & = \eta_j^{c_j}(s)\omega''_j, \quad j = 2, \ldots, t, \end{cases} \tag{4}$$

where $\omega''_j = \omega'_j/\omega_j^r$ is yet another fixed c_jth root of unity (assuming r is fixed).

A solution (n, s) to the equation $\mu^{\ell n} = \eta^\ell(s)$ implies that (n, s) is a solution to the equation $\mu^n = \omega\eta(s)$ for some ℓth root of unity ω. Conversely, any such solution is a solution to $\mu^{\ell n} = \eta^\ell(s)$. Therefore, to satisfy the first equation, we must have $\mu^n = \omega\eta(s)$ for some ω an ℓth root of unity. So assume (n, s) is a solution to the first equation. Then $\mu^{c_j n} = (\omega\eta_j)^{c_j} = \omega^{c_j}\eta^{c_j}(s)$. We conclude that (n, s) is a solution to (4) if and only if (n, s) is a solution to $\mu^n = \omega\eta(s)$ with ω an ℓth root of unity such that $\omega^{c_j} = \omega''_j$ for each $j = 2, \ldots, t$. We compute the set S_r of ℓth roots of unity ω which satisfy $\omega^{c_j} = \omega''_j$ for all j. Then the set of solutions (n, s) to (4) with $n = r \mod \ell$ is exactly the union of the solutions to the equations $\mu^n = \omega\eta(s)$, $\omega \in S_r$.

We claim that the characteristic sequences of the union of the solutions to the equations $\mu^n = \omega\eta(s)$, $w \in S_r$ can be expressed by an arc-hitting model. Indeed, the range of the function η is a finite union of arcs on the unit circle: indeed, we have that $|\eta(s)| = 1$ as this holds for γ. The domain of η might not be an interval, but might have at most finitely many points of discontinuity, but η maps each of its domain's connected components to an arc on the unit circle by continuity. The union of the unions of arcs given by $\omega\eta$, $\omega \in S_r$, gives a set $\mathcal{I}$ of arcs on the unit circle, such that $\mu^n \in \mathcal{I}$ if and only if $\mu^n = \omega\eta(s)$ for some $\omega \in S_r$. We note that the obtained arc-hitting model has angle $\arg(\lambda_1)/\ell$.

6 PUTTING HUMPTY TOGETHER AGAIN

In this section, we prove the main result of this paper, Theorem 3.1. From Theorem 4.1 and Theorem 5.1 we know that for a single semialgebraic target $\mathcal{T}$ that is either contained inside a 3D subspace or has intrinsic dimension (at most) 1, there exists computable $L > 0$ such that $\mathcal{Z}(\mathcal{T})$ is equal to interleaving of $\mathcal{Z}_0(\mathcal{T}), \ldots, \mathcal{Z}_{L-1}(\mathcal{T})$ where each $\mathcal{Z}_r(\mathcal{T}) = \{n \mid M^{nL+r}x \in \mathcal{T}\}$ can be represented by an arc-hitting model. Let us consider the situation where we have multiple targets $\mathcal{T}_1, \ldots, \mathcal{T}_m$ that are either contained in a subspace of dimension 3 or have intrinsic dimension (at most) 1. Suppose for $1 \le i \le m$, $\mathcal{Z}(\mathcal{T}_i)$ can be written as an interleaving of L_i sets, each of which can be represented by an arc-hitting model. Observe that if a set can be represented by L_i arc-hitting models, then for any positive integer multiple L of L_i, it can be represented using L arc-hitting models. Hence by taking the least common multiple of $L_1, \ldots, L_m$, we can assume that $L_i = L_j = L$ for every i, j. That is, we can assume that the number of subsequences we need to consider is equal to L for all targets $\mathcal{T}_1, \ldots, \mathcal{T}_m$.

From Section 2.5 we already know that for $0 \le r \le L - 1$ and $1 \le i \le m$, the set $\mathcal{Z}_r(\mathcal{T}_i)$ can be represented by an arc-hitting model and hence is effectively almost-periodic. Next we show that the overall word $w(O, \pi)$ with respect to targets $\mathcal{T}_1, \ldots, \mathcal{T}_m$, which is obtained by aggregating the sets $\mathcal{Z}_r(\mathcal{T}_i)$, is also effectively almost periodic.

Theorem 6.1. *Let (M, x) be a linear dynamical system such that $\mathcal{Z}_r(\mathcal{T}_1), \ldots, \mathcal{Z}_r(\mathcal{T}_m)$, for $r \in \{0, \ldots, L-1\}$, are represented by arc-hitting models. Let $w = w(\mathcal{O}, \pi) \in \left(2^{\{\mathcal{T}_1, \ldots, \mathcal{T}_m\}}\right)^{\mathbb{N}}$ be the characteristic word of the LDS with respect to the m targets. w is effectively almost periodic.*

Proof. Let $\lambda_{i,r}, N_{i,r}, F_{i,r}, I_{i,r}$ be the parameters of the arc-hitting model corresponding to $\mathcal{Z}_r(\mathcal{T}_i)$. We will use these arc-hitting models in the following way. Let $n = qL + r$, $0 \leq r < L$ be larger than $\max\{N_{i,r} : 1 \leq i \leq m, 0 \leq r < L\}$. We have that for each i, by the definition of arc-hitting models, $\mathcal{T}_i \in w[n]$ if and only if $\lambda_{i,r}^q \in I_{i,r}$. In other words, $\mathcal{T}_i \in w[n]$ if and only if $\lambda_{i,r}^{\lfloor \frac{n}{L} \rfloor} \in I_{i,r}$.

Next we compute large enough N such that for all $n > N$, $1 \leq i \leq m$ and $0 \leq r < L$,

- $\lfloor \frac{n}{L} \rfloor > N_{i.r}$, and
- $\lambda_{i,r}^{\lfloor \frac{n}{L} \rfloor}$ is not an endpoint of $I_{i,r}$. To see that this is possible, observe that for every i and r, the set $I_{i,r}$ has a finite number of endpoints. Since $\lambda_{i,r}$ by definition cannot be a root of unity, for every point $p \in \mathbb{T}$ one can compute N_p such that for all $n > N_p$, $\lambda^n \neq p$. Hence for all n larger than $\max\{N_p : p \text{ is an endpoint of } I_{i.r}\}$, λ^n is not an enpoint of $I_{i,r}$.

The first condition means that for $n > N$, $n = qL + r$, in order to determine whether $\mathcal{T}_i \in w[n]$ we only need to consider $\lambda_{i,r}$ and $I_{i,r}$ and ignore the finite set $F_{i,r}$ of exceptions. The second condition is useful for determining whether $\mathcal{T}_i \notin w[n]$: for $n > N$, $\mathcal{T}_i \notin w[n]$ if and only if $\lambda_{i,r}^q \notin I_{i,r}$, which, by the second condition is equivalent to $\lambda_{i,r}^q \in \mathrm{Int}(\mathbb{T} \setminus I_{i,r})$ (the interior of $\mathbb{T} \setminus I_{i,r}$). But crucially, $\mathrm{Int}(\mathbb{T} \setminus I_{i,r})$ is also a finite union of open arcs, and hence for $n > N$, $\mathcal{T}_i \in w[n]$ (or $\mathcal{T}_i \notin w[n]$) if and only if $\lambda_{i,r}$ is in a certain open semialgebraic subset of $\mathbb{T}$.

Let $\mathbb{T}^{L \cdot m}$ denote the $L \cdot m$-dimensional torus (with one coordinate per arc-hitting model and semialgebraic target). We define $\Lambda = (\lambda_{1,0}, \ldots, \lambda_{1,L-1}, \ldots, \lambda_{m,0}, \ldots, \lambda_{m,L-1}) \in \mathbb{T}^{L \cdot m}$ and write $\Lambda^n = (\lambda_{1,0}^n, \ldots, \lambda_{1,L-1}^n, \ldots, \lambda_{m,0}^n, \ldots, \lambda_{m,L-1}^n)$. Arc-hitting models describe the structure of $\mathcal{Z}_r(\mathcal{T}_i)$ in terms of powers of $\lambda_{i,r}$ and the open subset $I_{i,r}$ of $\mathbb{T}$. Next, we show how to describe the structure of w in terms of powers of Λ and semialgebraic open subsets of $\mathbb{T}^{L \cdot m}$.

Let $n = qL + r > N$, with $0 \leq r < L$.

- For every target $\mathcal{T}_i$, by definition of the arc-hitting model $\mathcal{T}_i \in w[n]$ if and only if $\lambda_{i,r}^q \in I_{i,r}$. Let $O_{i,r}$ be the preimage of $I_{i,r}$ under the projection map $\mathbb{T}^{L \cdot m} \to \mathbb{T}$ onto the coordinate (i, r). We have that $O_{i,r}$ is open and $\mathcal{T}_i \in w[n]$ if and only if $\Lambda^q \in O_{i,r}$.
- Similarly, for every target $\mathcal{T}_i$, let $O'_{i,r}$ be the preimage of $\mathrm{Int}(\mathbb{T} \setminus I_{i,r})$ in $\mathbb{T}^{L \cdot m}$ under the projection map onto (i, r). Recall that $\mathrm{Int}(\mathbb{T} \setminus I_{i,r})$ is the open set that is used to characterize the times when the orbit is not in $\mathcal{T}_i$. We have that $O'_{i,r}$ is open and $T_i \notin w[n]$ if and only if $\Lambda^q \in O'_{i,r}$, for all $n > N$.
- Next, let $\ell \in 2^{\{\mathcal{T}_1, \ldots, \mathcal{T}_m\}}$ be a letter. For example, suppose $\ell = \{\mathcal{T}_1, \mathcal{T}_3\}$, i.e. ℓ describes the set $(\mathcal{T}_1 \cup \mathcal{T}_3) \setminus \mathcal{T}_2$, assuming there are three targets in total. Then $w[n] = \ell$ if and only if $\Lambda^q \in O_{1,r}$ and $\Lambda^q \in O'_{2,r}$ and $\Lambda^q \in O_{3,r}$, which is equivalent to $\Lambda^q \in O_{1,r} \cap O'_{2,r} \cap O_{3,r} =: O_{\ell,r}$. Observing that $O_{\ell,r}$ is open, we conclude that for any letter ℓ, we can define an open subset $O_{\ell,r}$ (that is obtained by taking intersections and unions of the sets $O_{i,r}$ and $O'_{i,r}$) such that $w[n] = \ell$ if and only if $\Lambda^q \in O_{\ell,r}$.
- Finally, let $u = u_0 \cdots u_k$ be a finite word over $2^{\{\mathcal{T}_1, \ldots, \mathcal{T}_m\}}$. We characterize when u occurs at position n of w, that is for $0 \leq j \leq k$, $w[n + j] = u_j$. Let us consider a single equality $w[n + j] = u_j$. By the preceding analysis, this is equivalent to $\Lambda^{\lfloor \frac{n+j}{L} \rfloor} \in O_{u_j, n+j \bmod L}$. Observe that $\Lambda^{\lfloor \frac{n+j}{L} \rfloor} = \Lambda^{q+\delta_j} = \Lambda^q \Lambda^{\delta_j}$ for some $\delta_j \geq 0$. Therefore, $w[n + j] = u_j$ if and only if $\Lambda^q \in \Lambda^{-\delta_j} O_{u_j, n+j \bmod L}$, where multiplication is performed pointwise. Hence we obtain that

the word u occurs at position n if and only if $\Lambda^q \in O_{u,r}$ where

$$O_{u,r} = \bigcap_{0 \le j \le k} \Lambda^{-\delta_j} O_{u_j, n+j \bmod L}$$

is an open semialgebraic subset of $\mathbb{T}^{L \cdot m}$.

We now move onto proving effective almost periodicity. To this end, given a finite word u, we need to show how to compute a bound p_u such that either u does not occur in $w[p_u, \infty)$, or it occurs within every contiguous subword of w of length p_u. From the analysis above we can compute the open sets $O_{u,0}, \ldots, O_{u,L-1}$. Suppose all of these sets are empty. In this case, for $n > N$, the word u cannot occur in position $w[n]$. Hence we can choose $p_u = N$. Now suppose at least one of $O_{u,0}, \ldots, O_{u,L-1}$ is non-empty. Below we show how to compute an effective upper bound on the distance between consecutive occurrences.

Let $W = \{\Lambda^n : n \in \mathbb{N}\}$, where $\Lambda \in \mathbb{T}^{L \cdot m}$ is defined as above. First, compute an effective representation of the topological closure $\overline{W}$ of W.[12] The closure of W, unlike W itself, is very well-understood and semialgebraic; see [Ouaknine and Worrell 2014a, Appendix A] for how to compute a representation for it. It is also the case that, by Kronecker's theorem, the sequence $(\Lambda^n)_{n \in \mathbb{N}}$ is dense in $\overline{W}$ [Ouaknine and Worrell 2014a, Theorem 5].

Next, let $O_u = \bigcup_{r=0}^{L-1} O_{u,r}$. If $O_u \cap \overline{W}$ is empty (observe that this can be effectively checked as O_u and $\overline{W}$ are both semialgebraic), then Λ^n is never in O_u and hence for $n > N$, the word u cannot occur at position n of w. Therefore, we can once again choose $p_u = N$. It only remains to consider the case where $O_u \cap \overline{W}$ is non-empty. We will prove that in this case, the word u occurs infinitely often in w and show how to compute p_u. Wlog assume that $O_{u,0} \cap \overline{W} := O \neq \emptyset$, and observe that O must be open.

Recall that for large enough q (i.e. $qL > N$), $\Lambda^q \in O_{u,0}$ if and only if the word u occurs at position qL. By density of $(\Lambda^n)_{n \in \mathbb{N}}$ in $\overline{W}$ and openness of O, the sequence $(\Lambda^n)_{n \in \mathbb{N}}$ will visit O infinitely often. Hence the word u will occur at a position qL of w for infinitely many. To compute the bound on p_u on the gap between consecutive occurrences of u in w, we will proceed as follows. Consider the sequence $\langle O, \Lambda^{-1}O, \Lambda^{-2}O, \ldots \rangle$. By density of $(\Lambda^n)_{n \in \mathbb{N}}$ in $\overline{W}$, from any point in $\overline{W}$ one can reach O in finitely many steps under multiplication by Λ. Hence $\bigcup_{j=0}^{\infty} \Lambda^{-j} O = \overline{W}$. Notice that $\overline{W}$ is a closed, bounded set, and thus compact[13]. Therefore, there must exist a finite open subcover, i.e. $\bigcup_{j=0}^{M} \Lambda^{-j} O = \overline{W}$ for some $M > 0$. The value of M can be determined by guess and check: observe that for any prefix of $\langle O, \Lambda^{-1}O, \Lambda^{-2}O, \ldots \rangle$, whether it is a cover can be determined by manipulating the semialgebraic sets in the prefix and the semialgebraic set $\overline{W}$. In the end, we have that $(\Lambda^n)_{n \in \mathbb{N}}$ visits O within every M steps and hence the word u must occur at positions $(q_i L)_{i \in \mathbb{N}}$ where $q_{i+1} - q_i < M$. Therefore, u must occur within every window of size $L \cdot M$ in $w[N, \infty]$ and we can choose $p_u = \max\{N, L \cdot M\}$. □

Together with Theorem 1.1, Theorem 6.1 completes the proof of the main result of this paper.

ACKNOWLEDGMENTS

This work was funded by DFG grant 389792660 as part of TRR 248 (see https://perspicuous-computing.science). Joël Ouaknine was supported by ERC grant AVS-ISS (648701), and is also affiliated with Keble College, Oxford as emmy.network Fellow. James Worrell was supported by EPSRC Fellowship EP/N008197/1. Anton Varonka was supported by ERC Consolidator Grant ARTIST 101002685 and WWTF ICT19-018 grant ProbInG.

[12] The set $\overline{W}$ consists of W together with all of its limit points.

[13] A set S is compact if any open cover of S (a collection of open sets whose union contains S) has a finite subcover.

REFERENCES

Manindra Agrawal, S. Akshay, Blaise Genest, and P. S. Thiagarajan. 2015. Approximate Verification of the Symbolic Dynamics of Markov Chains. *J. ACM* 62, 1 (2015), 2:1–2:34. https://doi.org/10.1145/2629417

Shaull Almagor, Dmitry Chistikov, Joël Ouaknine, and James Worrell. 2018. O-Minimal Invariants for Linear Loops. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018 (LIPIcs, Vol. 107)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 114:1–114:14. https://doi.org/10.4230/LIPIcs.ICALP.2018.114

Shaull Almagor, Toghrul Karimov, Edon Kelmendi, Joël Ouaknine, and James Worrell. 2021a. Deciding ω-regular properties on linear recurrence sequences. *Proc. ACM Program. Lang.* 5, POPL (2021), 1–24. https://doi.org/10.1145/3434329

Shaull Almagor, Joël Ouaknine, and James Worrell. 2017. The Polytope-Collision Problem. In *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017 (LIPIcs, Vol. 80)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 24:1–24:14. https://doi.org/10.4230/LIPIcs.ICALP.2017.24

Shaull Almagor, Joël Ouaknine, and James Worrell. 2019. The Semialgebraic Orbit Problem. In *36th International Symposium on Theoretical Aspects of Computer Science, STACS 2019 (LIPIcs, Vol. 126)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 6:1–6:15. https://doi.org/10.4230/LIPIcs.STACS.2019.6

Shaull Almagor, Joël Ouaknine, and James Worrell. 2021b. First-Order Orbit Queries. *Theory Comput. Syst.* 65, 4 (2021), 638–661. https://doi.org/10.1007/s00224-020-09976-7

Christel Baier, Florian Funke, Simon Jantsch, Toghrul Karimov, Engel Lefaucheux, Florian Luca, Joël Ouaknine, David Purser, Markus A. Whiteland, and James Worrell. 2021. The Orbit Problem for Parametric Linear Dynamical Systems. , 28:1–28:17 pages. https://doi.org/10.4230/LIPIcs.CONCUR.2021.28 Extended version with proofs https://arxiv.org/abs/2104.10634.

Amir M. Ben-Amram, Jesús J. Doménech, and Samir Genaim. 2019. Multiphase-Linear Ranking Functions and Their Relation to Recurrent Sets. In *Static Analysis - 26th International Symposium, SAS 2019 (Lecture Notes in Computer Science, Vol. 11822)*. Springer, 459–480. https://doi.org/10.1007/978-3-030-32304-2_22

Amir M. Ben-Amram and Samir Genaim. 2013. On the linear ranking problem for integer linear-constraint loops. In *The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13*. ACM, 51–62. https://doi.org/10.1145/2429069.2429078

Amir M. Ben-Amram and Samir Genaim. 2014. Ranking Functions for Linear-Constraint Loops. *J. ACM* 61, 4 (2014), 26:1–26:55. https://doi.org/10.1145/2629488

Amir M. Ben-Amram and Samir Genaim. 2017. On Multiphase-Linear Ranking Functions. In *Computer Aided Verification - 29th International Conference, CAV 2017 (Lecture Notes in Computer Science, Vol. 10427)*. Springer, 601–620. https://doi.org/10.1007/978-3-319-63390-9_32

Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. 1998. *Real algebraic geometry*. Vol. 36. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03718-8

Bernard Boigelot. 2003. On iterating linear transformations over recognizable sets of integers. *Theor. Comput. Sci.* 309, 1-3 (2003), 413–468. https://doi.org/10.1016/S0304-3975(03)00314-1

Aaron R. Bradley, Zohar Manna, and Henny B. Sipma. 2005. Termination Analysis of Integer Linear Loops. In *CONCUR 2005 - Concurrency Theory, 16th International Conference, CONCUR 2005 (Lecture Notes in Computer Science, Vol. 3653)*. Springer, 488–502. https://doi.org/10.1007/11539452_37

Mark Braverman. 2006. Termination of Integer Linear Programs. In *Computer Aided Verification, 18th International Conference, CAV 2006 (Lecture Notes in Computer Science, Vol. 4144)*. Springer, 372–385. https://doi.org/10.1007/11817963_34

Jin-yi Cai, Richard J. Lipton, and Yechezkel Zalcstein. 2000. The Complexity of the A B C Problem. *SIAM J. Comput.* 29, 6 (2000), 1878–1888. https://doi.org/10.1137/S0097539794276853

Hong Yi Chen, Shaked Flur, and Supratik Mukhopadhyay. 2015. Termination proofs for linear simple loops. *Int. J. Softw. Tools Technol. Transf.* 17, 1 (2015), 47–57. https://doi.org/10.1007/s10009-013-0288-8

Ventsislav Chonev, Joël Ouaknine, and James Worrell. 2013. The orbit problem in higher dimensions. In *Symposium on Theory of Computing Conference, STOC'13*. ACM, 941–950. https://doi.org/10.1145/2488608.2488728

Ventsislav Chonev, Joël Ouaknine, and James Worrell. 2015. The Polyhedron-Hitting Problem. (2015), 940–956. https://doi.org/10.1137/1.9781611973730.64

Ventsislav Chonev, Joël Ouaknine, and James Worrell. 2016. On the Complexity of the Orbit Problem. *J. ACM* 63, 3 (2016), 23:1–23:18. https://doi.org/10.1145/2857050

Michael Colón, Sriram Sankaranarayanan, and Henny Sipma. 2003. Linear Invariant Generation Using Non-linear Constraint Solving. In *Computer Aided Verification, 15th International Conference, CAV 2003 (Lecture Notes in Computer Science, Vol. 2725)*. Springer, 420–432. https://doi.org/10.1007/978-3-540-45069-6_39

Michael Colón and Henny Sipma. 2001. Synthesis of Linear Ranking Functions. In *Tools and Algorithms for the Construction and Analysis of Systems, 7th International Conference, TACAS 2001 (Lecture Notes in Computer Science, Vol. 2031)*. Springer, 67–81. https://doi.org/10.1007/3-540-45319-9_6

Byron Cook, Andreas Podelski, and Andrey Rybalchenko. 2006a. Termination proofs for systems code. In *Proceedings of the ACM SIGPLAN 2006 Conference on Programming Language Design and Implementation*. ACM, 415–426. https://doi.org/10.1145/1133981.1134029

Byron Cook, Andreas Podelski, and Andrey Rybalchenko. 2006b. Terminator: Beyond Safety. In *Computer Aided Verification, 18th International Conference, CAV 2006 (Lecture Notes in Computer Science, Vol. 4144)*. Springer, 415–418. https://doi.org/10.1007/11817963_37

Patrick Cousot. 2005. Proving Program Invariance and Termination by Parametric Abstraction, Lagrangian Relaxation and Semidefinite Programming. In *Verification, Model Checking, and Abstract Interpretation, 6th International Conference, VMCAI 2005 (Lecture Notes in Computer Science, Vol. 3385)*. Springer, 1–24. https://doi.org/10.1007/978-3-540-30579-8_1

Patrick Cousot and Nicolas Halbwachs. 1978. Automatic Discovery of Linear Restraints Among Variables of a Program. In *Conference Record of the Fifth Annual ACM Symposium on Principles of Programming Languages 1978*. ACM Press, 84–96. https://doi.org/10.1145/512760.512770

Graham Everest, Alfred J. van der Poorten, Igor E. Shparlinski, and Thomas Ward. 2003. *Recurrence Sequences*. Mathematical surveys and monographs, Vol. 104. American Mathematical Society. http://www.ams.org/bookstore?fn=20&arg1=survseries&item=SURV-104

Nathanaël Fijalkow, Pierre Ohlmann, Joël Ouaknine, Amaury Pouly, and James Worrell. 2019. Complete Semialgebraic Invariant Synthesis for the Kannan-Lipton Orbit Problem. *Theory Comput. Syst.* 63, 5, 1027–1048. https://doi.org/10.1007/s00224-019-09913-3

Susanne Graf and Hassen Saïdi. 1997. Construction of Abstract State Graphs with PVS. In *Computer Aided Verification, 9th International Conference, CAV '97 (Lecture Notes in Computer Science, Vol. 1254)*. Springer, 72–83. https://doi.org/10.1007/3-540-63166-6_10

Ashutosh Gupta, Thomas A. Henzinger, Rupak Majumdar, Andrey Rybalchenko, and Ru-Gang Xu. 2008. Proving non-termination. In *Proceedings of the 35th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2008*. ACM, 147–158. https://doi.org/10.1145/1328438.1328459

Vesa Halava, Tero Harju, and Mika Hirvensalo. 2006. Positivity of second order linear recurrent sequences. *Discret. Appl. Math.* 154, 3 (2006), 447–451. https://doi.org/10.1016/j.dam.2005.10.009

Georges Hansel. 1985. A Simple Proof of the Skolem-Mahler-Lech Theorem. In *Automata, Languages and Programming, 12th Colloquium (Lecture Notes in Computer Science, Vol. 194)*. Springer, 244–249. https://doi.org/10.1007/BFb0015749

Michael A. Harrison. 1969. *Lectures on Linear Sequential Machines*. Academic Press, New York.

Mehran Hosseini, Joël Ouaknine, and James Worrell. 2019. Termination of Linear Loops over the Integers. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019 (LIPIcs, Vol. 132)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 118:1–118:13. https://doi.org/10.4230/LIPIcs.ICALP.2019.118

Bertrand Jeannet, Peter Schrammel, and Sriram Sankaranarayanan. 2014. Abstract acceleration of general linear loops. In *The 41st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '14*. ACM, 529–540. https://doi.org/10.1145/2535838.2535843

Ranjit Jhala, Andreas Podelski, and Andrey Rybalchenko. 2018. Predicate Abstraction for Program Verification. In *Handbook of Model Checking*. Springer, 447–491. https://doi.org/10.1007/978-3-319-10575-8_15

Ravindran Kannan and Richard J. Lipton. 1980. The Orbit Problem is Decidable. In *Proceedings of the 12th Annual ACM Symposium on Theory of Computing 1980*. ACM, 252–261. https://doi.org/10.1145/800141.804673

Ravindran Kannan and Richard J. Lipton. 1986. Polynomial-time algorithm for the orbit problem. *J. ACM* 33, 4 (1986), 808–821. https://doi.org/10.1145/6490.6496

Toghrul Karimov, Joël Ouaknine, and James Worrell. 2020. On LTL Model Checking for Low-Dimensional Discrete Linear Dynamical Systems. In *45th International Symposium on Mathematical Foundations of Computer Science, MFCS 2020 (LIPIcs, Vol. 170)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 54:1–54:14. https://doi.org/10.4230/LIPIcs.MFCS.2020.54

Manuel Kauers and Peter Paule. 2011. *The Concrete Tetrahedron - Symbolic Sums, Recurrence Equations, Generating Functions, Asymptotic Estimates*. Springer. https://doi.org/10.1007/978-3-7091-0445-3

Zachary Kincaid, Jason Breck, John Cyphert, and Thomas W. Reps. 2019. Closed forms for numerical loops. *Proc. ACM Program. Lang.* 3, POPL (2019), 55:1–55:29. https://doi.org/10.1145/3290368

Zachary Kincaid, John Cyphert, Jason Breck, and Thomas W. Reps. 2018. Non-linear reasoning for invariant synthesis. *Proc. ACM Program. Lang.* 2, POPL (2018), 54:1–54:33. https://doi.org/10.1145/3158142

Vichian Laohakosol and Pinthira Tangsupphathawat. 2009. Positivity of third order linear recurrence sequences. *Discret. Appl. Math.* 157, 15 (2009), 3239–3248. https://doi.org/10.1016/j.dam.2009.06.021

Engel Lefaucheux, Joël Ouaknine, David Purser, and James Worrell. 2021. Porous Invariants. In *Computer Aided Verification - 33rd International Conference, CAV 2021 (Lecture Notes in Computer Science, Vol. 12760)*. Springer, 172–194. https://doi.org/10.1007/978-3-030-81688-9_8

Martin Leucker and Christian Schallhart. 2009. A brief account of runtime verification. *J. Log. Algebraic Methods Program.* 78, 5 (2009), 293–303. https://doi.org/10.1016/j.jlap.2008.08.004

Nicolas Markey and Philippe Schnoebelen. 2003. Model Checking a Path. In *CONCUR 2003 - Concurrency Theory, 14th International Conference (Lecture Notes in Computer Science, Vol. 2761)*. Springer, 248–262. https://doi.org/10.1007/978-3-540-45187-7_17

David W. Masser. 1988. *Linear relations on algebraic groups.* Cambridge University Press, 248–262. https://doi.org/10.1017/CBO9780511897184.016

Maurice Mignotte. 1982. *Some Useful Bounds.* Springer Vienna, Vienna, 259–263. https://doi.org/10.1007/978-3-7091-3406-1_16

Andrei A. Muchnik, Alexei L. Semenov, and Maxim Ushakov. 2003. Almost periodic sequences. *Theor. Comput. Sci.* 304, 1-3 (2003), 1–33. https://doi.org/10.1016/S0304-3975(02)00847-2

Joël Ouaknine, João Sousa Pinto, and James Worrell. 2017. On the Polytope Escape Problem for Continuous Linear Dynamical Systems. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control, HSCC 2017.* ACM, 11–17. https://doi.org/10.1145/3049797.3049798

Joël Ouaknine and James Worrell. 2014a. On the Positivity Problem for Simple Linear Recurrence Sequences,. In *Automata, Languages, and Programming.* Springer Berlin Heidelberg, Berlin, Heidelberg, 318–329. https://doi.org/10.1007/978-3-662-43951-7_27 Extended version with proofs https://arxiv.org/abs/1309.1550.

Joël Ouaknine and James Worrell. 2014b. Positivity Problems for Low-Order Linear Recurrence Sequences. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014.* SIAM, 366–379. https://doi.org/10.1137/1.9781611973402.27

Joël Ouaknine and James Worrell. 2014c. Ultimate Positivity is Decidable for Simple Linear Recurrence Sequences. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014 (Lecture Notes in Computer Science, Vol. 8573)*. Springer, 330–341. https://doi.org/10.1007/978-3-662-43951-7_28

Joël Ouaknine and James Worrell. 2015. On linear recurrence sequences and loop termination. *ACM SIGLOG News* 2, 2 (2015), 4–13. https://dl.acm.org/citation.cfm?id=2766191

Andreas Podelski and Andrey Rybalchenko. 2004a. A Complete Method for the Synthesis of Linear Ranking Functions. In *Verification, Model Checking, and Abstract Interpretation, 5th International Conference, VMCAI (Lecture Notes in Computer Science, Vol. 2937)*. Springer, 239–251. https://doi.org/10.1007/978-3-540-24622-0_20

Andreas Podelski and Andrey Rybalchenko. 2004b. Transition Invariants. In *19th IEEE Symposium on Logic in Computer Science (LICS 2004)*. IEEE Computer Society, 32–41. https://doi.org/10.1109/LICS.2004.1319598

Enric Rodríguez-Carbonell and Deepak Kapur. 2004. An Abstract Interpretation Approach for Automatic Generation of Polynomial Invariants. In *Static Analysis, 11th International Symposium, SAS 2004 (Lecture Notes in Computer Science, Vol. 3148)*. Springer, 280–295. https://doi.org/10.1007/978-3-540-27864-1_21

Enric Rodríguez-Carbonell and Deepak Kapur. 2007. Generating all polynomial invariants in simple loops. *J. Symb. Comput.* 42, 4 (2007), 443–476. https://doi.org/10.1016/j.jsc.2007.01.002

Aleksei Lvovich Semenov. 1984. Logical theories of one-place functions on the set of natural numbers. *Mathematics of the USSR-Izvestiya* 22, 3 (1984), 587.

Sergey P. Tarasov and Mikhail N. Vyalyi. 2011. Orbits of Linear Maps and Regular Languages. In *Computer Science - Theory and Applications - 6th International Computer Science Symposium in Russia, CSR 2011 (Lecture Notes in Computer Science, Vol. 6651)*. Springer, 305–316. https://doi.org/10.1007/978-3-642-20712-9_24

Robert Tijdeman, Maurice Mignotte, and Tarlok Nath Shorey. 1984. The distance between terms of an algebraic recurrence sequence. *Journal für die reine und angewandte Mathematik* 349 (1984), 63–76. http://eudml.org/doc/152622

Ashish Tiwari. 2004. Termination of Linear Programs. In *Computer Aided Verification, 16th International Conference, CAV 2004 (Lecture Notes in Computer Science, Vol. 3114)*. Springer, 70–82. https://doi.org/10.1007/978-3-540-27813-9_6