Decidability of Approximate Skolem Problem and Applications to Logical Verification of Dynamical Properties of Markov Chains

M. Biscaia* D. Henriques*,†
P. Mateus*

* Department of Mathematics, Instituto Superior Técnico, Universidade de Lisboa
* SQIG - Instituto de Telecomunicações
† Department of Computer Science, Carnegie Mellon University

September 1, 2014

Abstract

When studying probabilistic dynamical systems, temporal logic has typically been used to reason about path properties. Recently, there has been some interest in reasoning about the dynamical evolution of state probabilities of these systems. In this paper we show that verifying linear temporal properties concerning the state evolution induced by a Markov chain is equivalent to the decidability of the Skolem problem – a long standing open problem in Number Theory. However, from a practical point of view, usually it is enough to check properties up to some acceptable error bound ϵ. We show that an approximate version of Skolem problem is decidable, and that it can be applied to verify, up to arbitrarily small ϵ, linear temporal properties of the state evolution induced by a Markov chain.

1 Introduction

Verification of properties in probabilistic deterministic systems is a critical area of research in the field of computer science. Currently, there are many tools that verify properties of systems modeled as Markov chains [14, 9, 12]. However most of the work is focused on verifying path-like specifications, that is, what proportion of possible executions of the system satisfy a given property [21, 3, 4]. These specifications are undoubtedly interesting since they are ubiquitous; however there are many interesting and intuitive properties that they can not express, in particular, considerations about the dynamical evolution of state probabilities are either convoluted or impossible to state in these frameworks, as pointed in e.g. [15, 13, 7, 1].

Recently Agrawal *et al.* [1, 2] developed work that builds upon well-known results in Probability Theory in order to reason with the dynamical system induced by a Markov chain. This work focuses on characterizing the behaviour of a Markov chain by the use of symbolic distributions that evolve dynamically and by considering error bounds for these evolutions. We work in the same setting, but follow a different approach. We display the deep connection between the exact verification of linear properties over the dynamical evolution of state probabilities and a famous open problem in Number Theory, the Skolem problem [18]. In fact, as remarked by Agrawal [2],"[the verification problem for linear time properties over the dynamical of state probabilities] seems to be strongly related to the long-standing open problem on linear recurrent sequences known as the Skolem problem".

The Skolem problem, originally formulated by Thoralf Skolem, was partially solved by himself using non-constructive techniques. More recently, thanks to the work in [16, 8], it has been reformulated as a decision problem which is only known to be decidable for low dimensions [18, 19].

Skolem Problem can be stated [18] as follows:

Problem 1. *Given $x, y \in \mathbb{Q}^m$, and $L \in \mathbb{Q}^{m^2}$, can we decide if the following statement is true:*

$$\exists n. x^T L^n y = 0$$

The problem itself can be seen whether the repeated mapping of a starting vector, x, through some linear system, L, and later projected upon another vector, y, ever reaches 0. In our probabilistic setting, we have a similar problem, a reachability query, where we ask whether, from an initial distribution (the analogue of x), a Markovian evolution (the analogue of L^i) ever reaches a point where a combination of the probabilities of states (the analogue of y) is exactly 0. It is easy to show that these two problems are equireducible. A more interesting question is whether the Skolem problem is easier than the verification of more complex linear time properties (other than simple reachability queries). We show that in fact, for linear time properties, the verification problem is as difficult as the Skolem problem.

We also pursue the subject in a different perspective. If the verification problem is as hard as the (open) Skolem problem, can we at least solve approximate versions of them? That is, if we relax Problem 1 in the following way:

Problem 2. *Given $x, y \in \mathbb{Q}^m$, and $L \in \mathbb{Q}^{m^2}$, there exists $d \in \mathbb{Q}$, such that for all $\epsilon \in]0; d[\cap \mathbb{Q}$ can we decide if the following statement is true:*

$$\exists n. -\epsilon < x^T L^n y < \epsilon,$$

can we present a procedure to solve the problem (and, respectively, one for an equivalent relaxation of the verification problem)?

Agrawal *et al.* [2] considered a similar problem from the verification of dynamical properties perspective. Our approach focuses more on the relation between the approximate Skolem problem and the approximate verification problem, solving both.

Our contributions

- We show the equireducibility between the Skolem problem and the verification of linear time properties for dynamical systems induced by Markov chains.
- We present decision procedures for approximate versions of both problems.

2 Preliminaries

In this section we introduce the main definitions and prior results used throughout the article.

2.1 Markov chains

Markov chains are a widely used formalism to model memoryless discrete probabilistic dynamic systems. Let Λ be a finite set of propositional symbols.

Definition 1. A *(state labelled) Markov chain* $\mathcal{M}$ is a tuple $\mathcal{M} = (S, M, L, \mu)$ over the propositional symbols Λ where:

- $S = \{s_1, \ldots, s_m\}$ is a finite set of states;
- M is a stochastic matrix of dimension m with entries in $\mathbb{Q}$, named the *transition matrix*. Intuitively, from s_i we can move towards s_j in one step with probability $M_{i,j}$;
- $L : S \mapsto \{0,1\}^\Lambda$ is the *labeling function*. Intuitively, it represents which predicates are true in each state;
- μ is a probability distribution over S, named the *initial distribution*.

A Markov chain induces a discrete dynamical system of finite probability distributions over states given by $\mu_0 = \mu, \mu_{i+1}^T = \mu_i^T M$. We are interested in verifying properties of this dynamical system. In order to do so, we need to express properties of probability distributions, and their evolution.

2.2 EPLTL as a logic for verification of probability distributions given by Markov chains

We consider the following logic, named EPLTL, already described and studied in [5, 17], which has the following syntax:

Definition 2. The well formed formulae in EPLTL over the propositional symbols Λ are given below in Backus-Naur notation:

$$\beta := p \;[]\; (\beta \wedge \beta) \;[]\; (\neg\beta) \qquad \text{(basic formulae)}$$

$$t := r \;[]\; (t + t) \;[]\; r \int \beta \qquad \text{(rational terms)}$$

$$\delta := \underbrace{\sim \delta \;[]\; \delta \cap \delta \;[]\; \mathsf{X}\delta \;[]\; \delta \mathsf{U} \delta}_{\text{temporal formulae}} \quad [] \quad \underbrace{t = t}_{\text{comparison formulae}} \qquad \text{(formulae)}$$

where $p \in \Lambda$ and $r \in \mathbb{Q}$.

The logic will enable us to reason about distributions over the propositional symbols Λ using terms of the form $\int \beta$ and comparison formulae; the linear temporal reasoning is performed using the outer propositional negation $\sim$, conjunction $\cap$, the *neXt* time connective, and the *Until* connective. The common abbreviations for *sometime in the Future* and *Globally* will also be considered.

The semantics considered in this article are designed specifically towards Markov chains.

Definition 3. The denotation of a rational term t in a Markov chain $\mathcal{M} = (S, M, L, \mu)$ in the instant $i \in \mathbb{N}$, denoted by $[\![t]\!]_{\mathcal{M},i}$, is defined inductively as:

- $[\![r]\!]_{\mathcal{M},i} = r$,
- $[\![r \int \beta]\!]_{\mathcal{M},i} = r\mu_i(\{s \in S : L(s) \Vdash^{\mathsf{PL}} \beta\})$,
- $[\![(t_1 + t_2)]\!]_{\mathcal{M},i} = [\![t_1]\!]_{\mathcal{M},i} + [\![t_2]\!]_{\mathcal{M},i}$,

where $\Vdash^{\mathsf{PL}}$ is the satisfaction relation in propositional logic. Finally the satisfaction relation between a Markov chain $\mathcal{M}$, an instant i and a formula δ can be defined.

Definition 4. The satisfaction relation between a Markov chain $\mathcal{M}$, an instant i, and a EPLTL formula δ is defined inductively:

- $\mathcal{M}, i \Vdash (t_1 = t_2)$ iff $[\![t_1]\!]_{\mathcal{M},i} = [\![t_2]\!]_{\mathcal{M},i}$,
- $\mathcal{M}, i \Vdash (\delta_1 \cap \delta_2)$ iff $\mathcal{M}, i \Vdash \delta_1$ and $\mathcal{M}, i \Vdash \delta_2$,
- $\mathcal{M}, i \Vdash (\sim \delta)$ iff $\mathcal{M}, i \nVdash \delta$,
- $\mathcal{M}, i \Vdash (\mathsf{X}\delta)$ iff $\mathcal{M}, i + 1 \Vdash \delta$,
- $\mathcal{M}, i \Vdash (\delta_1 \mathsf{U} \delta_2)$ iff there exists $j \geq i$ such that $\mathcal{M}, j \Vdash \delta_2$ and $\mathcal{M}, k \Vdash \delta_1$, for all $i \leq k < j$.

We say that a Markov chain $\mathcal{M}$ is a *model* of an EPLTL formula δ if $\mathcal{M}, 0 \Vdash \delta$.

Remark 1. Even though the classical model-checking logics like PCTL [10] are quite well suited to deal with properties about *probabilities of evolutions*, they are unable to deal with properties about the *evolution of probabilities*; for instance, consider the Markov chain in Figure 1. The assertion A "the probability of reaching s_2 sometime in the future is $\frac{1}{2}$" is checked by considering the *probability of the set of paths* passing by the state s_2; we are measuring sets of evolutions. However, the statement B "there is an instant of time such that the probability of being in s_2 is $\frac{1}{2}$" requires, with absolute certainty, that there exists an instant in time such that s_2 holds with probability $\frac{1}{2}$.

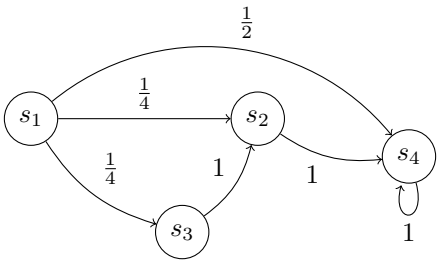

Figure 1: An example of the problem described. For $\mu_0 = (1, 0, 0, 0)$, there is no evolution such that s_2 holds with probability $\frac{1}{2}$, but the set of all evolutions such that s_2 holds has probability $\frac{1}{2}$.

Remark 2. Finally, while we allow any general labeling function on our Markov chains, we will just consider, without loss of generality, that the labeling function $L : S \mapsto \{0, 1\}^S$ maps a state s to the valuation that is only true on s. We can easily rewrite EPLTL formulae under one of the labelings into equivalent formulae under the other. In order to shorten notation,henceforth we will represent a Markov chain $\mathcal{M} = (S, M, L, \mu)$ simply by (M, μ).

2.3 The Skolem Problem

We now describe a problem from Number Theory that surprisingly enough is the cornerstone for verification of EPLTL formulae in Markov chains.

The Skolem problem is usually stated as a decision problem over linear recurrence sequences. A linear recurrence sequence x_n of order k over the rationals is defined as:

$$\begin{aligned} &x_0 = c_0, \ldots, x_{k-1} = c_{k-1}, \\ &x_n = a_{k-1}x_{n-1} + a_{k-2}x_{n-2} + \ldots + a_0 x_{n-k}, \\ &\quad c_0, \ldots, c_{k-1}, a_0, \ldots, a_{k-1} \in \mathbb{Q} \end{aligned}$$

Thoralf Skolem investigated [20] whether one could characterize the set

$Z(\{x_n\}) = \{n \in \mathbb{N} : x_n = 0\}$, for each sequence x_n. Originally Skolem proved that the set $Z(\{x_n\})$ could be written as:

$$Z(\{x_n\}) = F \cup G_1 \cup G_2 \cup \ldots \cup G_j,$$
where F is a finite (possibly empty) set and
$G_1, \ldots G_j$ are arithmetical progressions.

Unfortunately the proof was non-constructive. Later on, it was proved that all coefficients of all the arithmetical progressions are effectively computable [8]. However, there is still no known means of producing F, or testing its emptiness, which leaves the original problem open. The Skolem problem can also be approached by considering matrices instead of linear recurrence sequences:

Given $x, y \in \mathbb{Q}^k, L \in \mathbb{Q}^{k^2}$ characterize the set $Z(x^T L^n y) = \{n \in \mathbb{N} : x^T L^n y = 0\}$

Both versions are equireducible and so we also know that $Z(x^T L^n y)$ can be written as the union of a finite set and a finite union of arithmetical progressions. Since the coefficients of the arithmetical progressions are computable, the characterization problem can be restated as the following decision problem:

Problem 3 (Skolem Problem)**.** *Given $x, y \in \mathbb{Q}^m$ and $L \in \mathbb{Q}^{m^2}$, decide if the following statement is true:*

$$\exists n . x^T L^n y = 0$$

This more modern restatement is the one followed by the literature [18, 19]; we will motivate this restatement with a simple example:

Example 1. Assume that the Skolem problem is decidable. Now, suppose that $Z(x^T L^n y) = F \cup G_{q\mathbf{k}+r}$, such that we do not know F, but we can compute the period of the arithmetical progression q, and the shift r. Then, we can compute F by repeatedly querying whether some subsequences of $x^T L^n y$ have any zeros.

- Check whether $Z(x^T (L^q)^n L^{r+1} y)$ is empty. If so, we know that there are no zeros of the form $q\mathbf{k} + r + 1$. Otherwise, compute the index of the first zero of this form, then consider the subsequence beginning at that position and reiterate the process until the Skolem oracle returns that there are no more zeroes. Repeat the procedure for each of the $q - 1$ residue classes except the one with infinitely many zeroes, $q\mathbf{k} + r$.

- This procedure will terminate, since there are only a finite number of non-periodic zeroes.

- If $Z(x^T L^n y)$ is described by more than one arithmetical progression, for instance with periods q_1, q_2, then one adapts the algorithm to be applied to the least common multiple amongst the periods.

We will denote the union of a finite set F and finitely many arithmetical progressions $p_1\mathbf{k}+r_1, \ldots, p_n\mathbf{k}+r_n$ by a pair (F, G), where $G = \{(p_1, r_1), \ldots, (p_n, r_n)\}$ (henceforth named *representation*) . Elements in F will be named *exceptional zeroes*. We say that $m \in \mathbb{N}$ *is represented* in (F, G) if either $m \in F$ or $m = p_i\mathbf{k} + \mathbf{r_i}$ for some $k \in \mathbb{N}$ and $(p_i, r_i) \in G$. Likewise, we say that a set S is *representable* if there exists a pair (F, G) such that $m \in S$ iff m is represented in (F, G). We state without proof the following fact: the Skolem problem is decidable iff for any rational matrix L, vectors x, y and representation (F, G), we can decide whether the set of elements represented by (F, G) is precisely the set $Z(x^T L^n y)$.

We note, furthermore, that the Skolem problem is equireducible to the following problem [18]:

Problem 4. *Given $x, y \in \mathbb{Q}^m$, $c \in \mathbb{Q}$ and $L \in \mathbb{Q}^{m^2}$, decide if the following statement is true:*

$$\exists n. x^T L^n y = c$$

We can now see why this problem is relevant for verification of EPLTL formulae. For example, checking whether $\mathcal{M} \Vdash \mathsf{F}(\int s_1 - \int s_2) = \frac{1}{2}$ can be seen to be an instance of the Skolem problem with μ as x, the Markov chain matrix as L, and y as the vector $(1, -1, 0, ..., 0)$ and $c = \frac{1}{2}$. One might entertain the thought that, due to L being a Markov chain, these instances of the Skolem problem might be easier than the general statement. However, that is not case, as we now show.

2.3.1 Skolem Problem over Markov chains

The reduction between the general Skolem problem and the case for Markov chains can be done in two simple steps; assume that we are given $x, y \in \mathbb{Q}^m$ and $L \in \mathbb{Q}^{m^2}$; then define $x', y' \in \mathbb{Q}^{2m+2}$ and $L' \in \mathbb{Q}^{(2m+2)^2}$ as follows:

- $x' \leftarrow (1, 0, \mathbf{0})$,
- $L' \leftarrow \begin{pmatrix} 0 & x \\ \mathbf{0} & L \end{pmatrix}^{\dagger}$,
- $y' \leftarrow (0, 0, y_1, -y_1, \ldots, y_m, -y_m)$

where $\mathbf{0}$ is a subvector or matrix of the appropriate size.

The transformation $(N)^{\dagger}$ applied to the $(m+1) \times (m+1)$ matrix consists in rewriting each entry $n_{i,j}$ as the submatrix $\begin{pmatrix} n^1_{i,j} & n^2_{i,j} \\ n^2_{i,j} & n^1_{i,j} \end{pmatrix}$, where $n^1_{i,j}, n^2_{i,j}$ are non-negative numbers such that $n_{i,j} = n^1_{i,j} - n^2_{i,j}$. It is easy to see that due to the change in the vector y', $x'^T L'^i y' = x^T L^i y$ holds. Now, we have a stochastic vector as the initial distribution, and a non negative matrix L'. In order to obtain a stochastic matrix, we will add an extra dimension wich allows us to normalize each line. Let $K = \max_i \{\sum_j L'_{i,j}\}$:

- $x'' \leftarrow (x', 0)$,
- $L'' \leftarrow \frac{1}{K} \begin{pmatrix} L'_{1,1} & ... & L'_{1,n} & K - \sum_j L'_{1,j} \\ ... & ... & ... & ... \\ L'_{n,1} & ... & L'_{n,n} & K - \sum_j L'_{n,j} \\ 0 & ... & 0 & K \end{pmatrix}$,
- $y'' \leftarrow (y', 0)$.

So, in fact, by considering stochastic matrices and an initial stochastic vector we do not obtain an easier problem. We can also consider the probabilistic version with equality to a constant (rather than equality to 0 only) by adapting the proof of the reduction used in Problem 4. The major issue now is the fact that the Skolem problem is open. Given two vectors x, y and a matrix L, there is no known algorithm to decide whether there exists an index n such that $x^T L^n y = 0$ (except for cases of dimension less or equal to 4). Moreover, while the Skolem problem is clearly an extremely specific case of a verification problem for EPLTL, we would like to know whether more complex EPLTL formulae will be even harder to verify.

3 Using Skolem problem as an oracle

We now show that the problem of verification of EPLTL formulae in Markov chains is equireducible to the Skolem problem. Without loss of generality, we will assume that any comparison term is of the form $t = 0$. We will show that given representations, (F_i, G_i), for the atomic equalities of rational terms $t_i = 0$, we can compute a representation for the whole formula. This process is possible due to sets describable as the union of a finite set and a finite union of arithmetical progressions are closed under EPLTL connectives.

Remark 3. Note that each rational term $\sum_{i=1}^{N} r_i \int s_i + c$ is an affine combination of probabilities of individual states; therefore we can represent this term as the vector t^* of $N + 1$ components $(r_1, \ldots, r_N, -c)$. In order to compute whether $t_i = 0$ holds after n iterations of a Markov chain $(M.\mu)$ we just need to compute whether $\mu^T M t^* = 0$. We will overload the notation by also denoting t^* by t.

Definition 5. The satisfaction set of $\delta \in$ EPLTL in a Markov chain $\mathcal{M} = (M, \mu)$, represented as I_δ, is given inductively as:

- $I_{t=0} = \{i \in \mathbb{N} : \mu^T M^i t = 0\}$,
- $I_{\sim\delta_1} = \overline{I_{\delta_1}}$,
- $I_{\delta_1 \cap \delta_2} = I_{\delta_1} \cap I_{\delta_2}$,
- $I_{\delta_1 \cup \delta_2} = I_{\delta_1} \cup I_{\delta_2}$,
- $I_{\mathsf{X}\delta_1} = \{i \in \mathbb{N} : i + 1 \in I_{\delta_1}\}$,

- $I_{\mathsf{F}\delta_1} = \{i \in \mathbb{N} : \exists j \text{ s.t. } j \geq i, j \in I_{\delta_1}\}$,
- $I_{\delta_1 \mathsf{U} \delta_2} = \{i \in \mathbb{N} : \exists j \text{ s.t. } j \in I_{\delta_2} \text{ and } \forall k \text{ s.t. } i \leq k < j, k \in I_{\delta_1}\}$.

The proof of the following lemma is a simple exercise in structural induction.

Lemma 1. Let $\mathcal{M} = (M, \mu)$ be a Markov chain and let I_δ be the satisfaction set of $\delta \in \mathsf{EPLTL}$ in $\mathcal{M}$; then $(M, \mu), i \Vdash \delta$ iff $i \in I_\delta$.

We will now show that given representations (F_i, G_i) for the basic terms $t_i = 0$ of a formula δ, then I_δ is the union of a finite set and arithmetical progressions, and we can also provide a representation (F_δ, G_δ) for I_δ.

Proposition 1. Let $(F_1, G_1), (F_2, G_2)$ be representations for $I_{\delta_1}, I_{\delta_2}$, formulae in EPLTL. Then there exist representations for:

$$I_{\sim\delta_1} \quad I_{\delta_1 \cup \delta_2} \quad I_{\delta_1 \cap \delta_2} \quad I_{\mathsf{X}\delta_1} \quad I_{\mathsf{F}\delta_1}.$$

Proof. The case of the **union** connective is trivial. If (F_1, G_1) is a representation of I_{δ_1} and (F_2, G_2) is a representation of I_{δ_2}, then a representation for $I_{\delta_1 \cup \delta_2}$ is simply $(F_1 \cup F_2, G_1 \cup G_2)$.

The case of the **intersection** is slightly more involved, relying on the fact that the intersection of two arithmetical progressions is still an arithmetical progression or the empty set. The resulting arithmetical progression can be computed using the Chinese Remainder Theorem[11, pp 873-876], even if the periods are not coprime. If $I_{\delta_1} = F_1 \cup G_{p_1\mathbf{k}+r_1} \cup \ldots \cup G_{p_m\mathbf{k}+r_m}$ and $I_{\delta_2} = F_2 \cup G_{q_1\mathbf{k}+s_1} \cup \ldots \cup G_{q_n\mathbf{k}+s_n}$, then their intersection can be written as:

$$I_{\delta_1} \cap I_{\delta_2} = (F_1 \cap F_2) \cup \bigcup_{j=1}^{n} (F_1 \cap G_{q_j\mathbf{k}+s_j}) \cup \bigcup_{i=1}^{m} (F_2 \cap G_{p_i\mathbf{k}+r_i}) \cup \bigcup_{i,j}^{n,m} (G_{q_i\mathbf{k}+s_i} \cap G_{p_j\mathbf{k}+r_j})$$

Now clearly the finite part can be computed, and since each intersection of arithmetical progressions is either the empty set or another arithmetical progression, then it is possible to obtain a representation for $I_{\delta_1 \cap \delta_2}$.

The case of the **complement** is naturally related with the intersection. The complement of an arithmetical progression $p\mathbf{k}+r$ can be seen to be the union of the finite set $\{0, 1, \ldots, r-1\}$ and $p-1$ arithmetical progressions $p\mathbf{k}+r+i, i \in \{1, 2, \ldots, p-1\}$. Moreover the complement of a finite set F is again the union of the finite set $\{0, \ldots, max(F)\} - F$ with the trivial arithmetical progression $\mathbf{k} + max(F) + 1$. Therefore, if $I_{\delta_1} = F \cup G_{p_1\mathbf{k}+r_1} \cup G_{p_n\mathbf{k}+r_n}$,

$$I_{\sim\delta_1} = \overline{I_{\delta_1}} = \overline{F} \cap \bigcap_{i}^{n} \overline{G_{p_i\mathbf{k}+r_i}},$$

and so, thanks to the results already proved about the intersection, we prove that $I_{\sim\delta_1}$ is still the union of a finite set and a finite number of arithmetical progressions. All the coefficients can be computed and so we can obtain a representation for $I_{\sim\delta_1}$.

Regarding the temporal connectives, the X **connective** is quite straightforward; assuming $(F_1, \{(p_1, r_1), \ldots, (p_n, r_n)\})$ is a representation for I_{δ_1}, a representation for $I_{\mathsf{X}\delta_1}$ is $(F_1', \{(p_1, r_1'), \ldots, (p_n, r_n')\})$, where F_1' is the set of predecessors (in the natural numbers) of F_1 and $r_i' = r_i - 1 \mod p_i$.

While the F **connective** can be obtained from the U connective, we present it here for completeness sake. Suppose we have a representation of $I_{\delta_1} = (F, G)$. Then there are three cases: either $G \neq \emptyset$, and in this case a representation for I_{δ_1} is for example $(\emptyset, \{(1, 0)\})$, or $G = \emptyset$; if it is the latter case, if $F \neq \emptyset$ then a representation for $I_{\mathsf{F}\delta_1}$ would be $(\{0, 1, \ldots, max(F)\}, \emptyset)$, and $(\emptyset, \emptyset)$, otherwise. □

We will finally extend the result to consider the U connective. The proof relies on computing the representation of $I_{\delta_1 \mathsf{U} \delta_2}$ from a suitable finite domain, which captures all the information required. We motivate the proof with an extremely simple example:

Example 2. Suppose that we are given representations for $I_{\delta_1} = (\emptyset, \{(3, 3), (2, 1)\})$ and $I_{\delta_2} = (\emptyset, \{(3, 3), (2, 1)\})$, so δ_1 is true at the indexes given by $3\mathbf{k} + 3$ and $2\mathbf{k} + 1$, while δ_2 is true at the indexes $6\mathbf{k} + 2$ and also at the exceptional index 1 as depicted on Figure 2. Now let $T = lcm(3, 2, 6) = 6$ and $K = max\{\emptyset \cup \{1\} \cup \{3, 1, 2\}\} = 3$ so that the pattern of δ_1 and δ_2 repeats with period $T = 6$ after the index $K = 3$ (as can be seen in the Figure).

We now wish to capture at which indexes is $\delta_1 \mathsf{U} \delta_2$ satisfied. In this case a representation (F, G) for $I_{\delta_1 \mathsf{U} \delta_2}$ would take $F = \{1, 2\}$, as both exceptional indexes have δ_2 as label. To build G we analyse the truth values of δ_1 and of δ_2 in the first sequence of $T = 6$ indexes after $K = 3$. This sequence will repeat because:

- its period is a multiple of all the periods involved (all arithmetic progressions will repeat with this period, some more than once),
- the indexes considered are large enough to make sure all arithmetical progressions are represented (since they all start at most at K),
- the indexes considered are large enough to make sure that none of the exceptional indexes are represented (since they appear only until K).

So, to identify G, we can just consider the labelling from $i = K + 1 = 4$ until $i = K + T = 3 + 6 = 9$. It is clear that, in this range, $\delta_1 \mathsf{U} \delta_2$ holds at indexes $5, 6, 7, 8$, so we can consider $G = \{(6, 5), (6, 6), (6, 7), (6, 8)\}$.

We start by showing that the labeling function used in the example is (eventually) periodic with period T, given by the least common multiple of all the periods of the arithmetical progressions, after an initial segment of irregularities with size at most K, given by the maximum of all the residues from the arithmetical progressions and the exceptional zeroes.

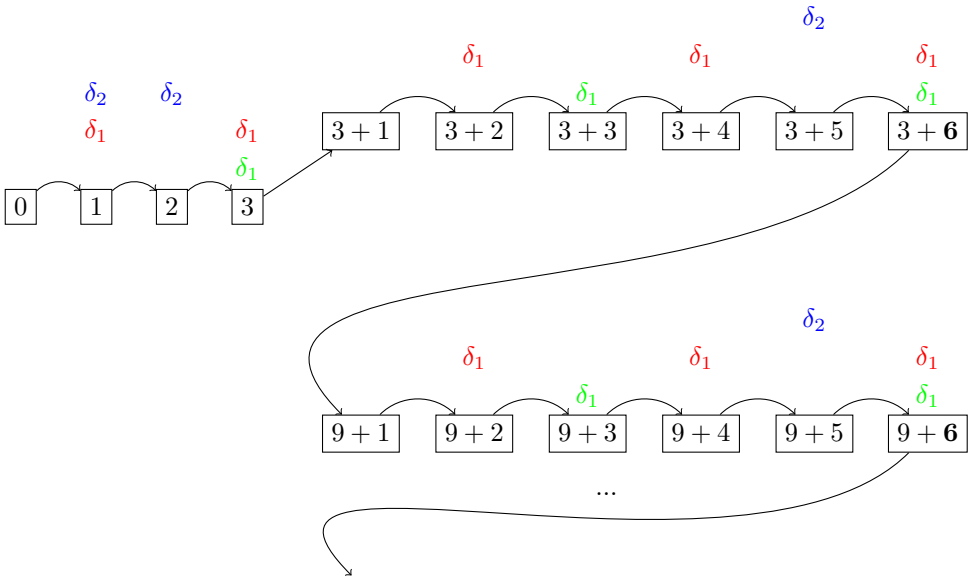

Figure 2: a labeling of each index assuming that the arithmetical progressions for δ_1 are $2\mathbf{k}+1$ and $3\mathbf{k}+3$ and the arithmetical progression for δ_2 is $6\mathbf{k}+2$. Furthermore, index 1 is labeled with δ_2 as I_{δ_2} has an exceptional zero. Note how any possible irregularities are discarded by setting K as large as necessary to consider all the exceptional zeroes and the starting point of all arithmetical progressions; note further that all arithmetical progressions repeat with period 6.

Lemma 2. Let $\delta_1, \delta_2 \in \mathsf{EPLTL}$. Given a Markov chain (M, μ), let I_{δ_i} be the satisfaction sets of δ_i and suppose that there exist representations (F_i, G_i) for them. Then $f : \mathbb{N} \to \mathcal{P}\{\delta_1, \delta_2\}$ defined by $\delta_i \in f(x)$ iff $x \in I_{\delta_i}$ is eventually periodic.

Proof. Assume that $G_1 = \{(p_1, r_1), \ldots, (p_n, r_n)\}$, $G_2 = \{(q_1, s_1), \ldots, (q_m, s_m)\}$. Let $K = \max(F_1 \cup F_2 \cup \{r_1, \ldots, r_n\} \cup \{s_1, \ldots, s_m\})$, $T = lcm(p_1, \ldots p_n, q_1, \ldots q_m)$.

Then for all natural numbers $n > K$, we need to prove that $f(n+T) = f(n)$. Suppose that $\delta_1 \in f(n)$; then n must be of the form $p_i\mathbf{k} + r_i$, since $n > K$. Then $n + T = p_i\mathbf{k} + r_i + T = q_i \frac{T}{p_i}\mathbf{k} + r_i$, which since $p_i | T$, implies that $\delta_1 \in f(n+T)$. The same argument can be applied to δ_2. □

Consider the following algorithm to compute a representation for $I_{\delta_1 \cup \delta_2}$ given representations for I_{δ_1} and I_{δ_2}, and the function defined as above:

Lemma 3. Let $\delta_1, \delta_2 \in \mathsf{EPLTL}$. Assume that $(F_1, G_1 = \{(p_1, r_1), \ldots (p_n, r_n)\})$ is a representation for I_{δ_1} and $(F_2, G_2 = \{(q_1, s_1), \ldots (q_n, s_n)\})$ is a representa-

Algorithm 1: UntilRepresentation computes a representation for the satisfaction set of a Until formula

Input: Representations $(F_1, G_1 = \{(p_1, r_1), \ldots, (p_n, r_n)\}), (F_2, G_2 = \{(q_1, s_1), \ldots, (q_n, s_n)\})$ for I_{δ_1}, I_{δ_2}

Output: A representation for $I_{\delta_1 \mathsf{U} \delta_2}$

$T \leftarrow lcm(p_1, \ldots, p_n, q_1, \ldots, q_n)$;

$K \leftarrow \max(\{r_1, \ldots, r_n\} \cup \{s_1, \ldots, s_n\} \cup F_1 \cup F_2)$;

$L \leftarrow \{(0, f(0)), \ldots (T + K, f(T + K))\}$ (as in Lemma 2) ;

foreach $(i, f(i)) \in L$ **do**

  if $\delta_2 \in f(i)$ **then**

    if $i \leq K$ **then**

      $F \leftarrow F \cup \{i\}$;

    else

      $G \leftarrow G \cup \{(T, i)\}$;

    end

  else if $\delta_1 \in f(i)$ **then**

    $j \leftarrow$ **Search for** *the least* j *s.t.* $(j, f(j)) \in L, j \geq i, \delta_2 \in f(j)$;

    if *Search succeeds* **then**

      if $\delta_1 \in f(k)$ *for every* $i \leq k < j$ **then**

        if $i \leq K$ **then**

          $F \leftarrow F \cup \{i\}$;

        else

          $G \leftarrow G \cup \{(T, i)\}$;

        end

      end

    else

      $j \leftarrow$

      Search for *the least* j *s.t.* $(j, f(j)) \in L, j \geq K + 1, \delta_2 \in f(j)$;

      if *Search succeeds* **then**

        if $\delta_1 \in f(k)$ *for every* $k \in \{i, \ldots, T + K, K + 1, \ldots, j - 1\}$ **then**

          $G \leftarrow G \cup \{(T, i)\}$;

        end

      end

    end

end

Return (F, G);

tion for I_{δ_2}. Then $(F, G = (T, a_1), \ldots (T, a_o))$ computed used using Algorithm 1 is a representation for $I_{\delta_1, \mathsf{U} \delta_2}$.

Proof. We note that Algorithm 1 halts for any input. We start by showing that if $n \in \mathbb{N}$ is represented in (F, G) then $n \in I_{\delta_1 \mathsf{U} \delta_2}$.

Suppose that $n \in F$. Then either $\delta_2 \in f(n)$ or $\delta_1 \in f(n)$ and there exists $n \leq m \leq T + K$ s.t $\delta_2 \in f(m)$. These are the only possible conditions in 1 where n could have been added to F.

- if $\delta_2 \in f(n)$, by the definition of f in Lemma 2, we know that $n \in I_{\delta_2}$, and as such $n \in I_{\delta_1 \mathsf{U} \delta_2}$.
- if $\delta_1 \in f(n)$ and $n \leq m \leq T + K$ s.t $\delta_2 \in f(m)$, we also know that all indexes l with $n \leq l < m$ are such that $\delta_1 \in f(l)$. Therefore, by the definition of f in Lemma 2 we obtain that $n \in I_{\delta_1}$, $l \in I_{\delta_1}$ and $m \in I_{\delta_2}$; as such $n \in I_{\delta_1 \mathsf{U} \delta_2}$.

Now suppose that $n = T\mathbf{k} + a_1$, that is the pair (T, a_1) was added to G; in any of the three possible commands where we might have added (T, a_1), we are guaranteed that $K < a_1 \leq T + K$.

- suppose that $\delta_2 \in f(a_1)$. Then, we automatically know that $\delta_2 \in f(a_1 + T\mathbf{k})$, as f is periodic after K, with period T. But if $\delta_2 \in f(a_1)$ then $a_1 \in I_{\delta_2}$, $a_1 + T\mathbf{k} \in I_{\delta_2}$ and therefore $n \in I_{\delta_1 \mathsf{U} \delta_2}$.
- otherwise suppose that $\delta_1 \in f(a_1)$ and that there exists an index m s.t $a_1 \leq m \leq T + K$ and $\delta_2 \in f(m)$. Again, we are assured that for any indexes l between a_1 and m, all of them are such that $\delta_1 \in f(l)$. Therefore $\delta_1 \in f(a_1 + T\mathbf{k})$, $\delta_1 \in f(l + T\mathbf{k})$, for $a_1 \leq l < m$, and $\delta_2 \in f(m + T\mathbf{k})$; so we can conclude that in fact $n \in I_{\delta_1 \mathsf{U} \delta_2}$.
- finally, the remaining possibility is the following: $\delta_1 \in f(a_1)$, and there exists $m \in \{K + 1, \ldots, a_1 - 1\}$ such that $\delta_2 \in f(m)$; furthermore, we also know that $\delta_1 \in f(l)$ for any $l \in \{a_1, \ldots T + K\} \cup \{K + 1, \ldots, m - 1\}$. But then, $\delta_1 \in f(a_1), \ldots, \delta_1 \in f(T + K), \delta_1 \in f(T + K + 1), \ldots \delta_1 \in f(T + K + m - 1)$ and $\delta_2 \in f(T + K + m)$; by using the periodicity of f, we obtain that in fact $a_1 + T\mathbf{k} \in I_{\delta_1 \mathsf{U} \delta_2}$.

Suppose that $n \in I_{\delta_1 \mathsf{U} \delta_2}$. We show that n is represented in (F, G). If $n \in I_{\delta_1 \mathsf{U} \delta_2}$, then there exists $m \in I_{\delta_2}$, with $m \geq n$ and for all $n \leq k < m$, $k \in I_{\delta_1}$. Then, we know that $\delta_1 \in f(k)$, with $n \leq k < m$ and $\delta_2 \in f(m)$. Assuming that $n > K$, we can use the periodicity of f to guarantee that there exists n' s. t. $K < n' \leq T + K$,$\delta_1 \in f(n')$. The same argument can be applied to the index m, obtaining $m' \in \{K + 1, \ldots, T + K\}$, $\delta_2 \in f(m')$. Afterwards we will always denote l' as the index obtained by the use of the periodicity of f applied to l in the range between $\{K + 1, \ldots T + K\}$.

Suppose that $m' = n' + l$, $l \geq 0$. Then, consider the indexes $n + 1, \ldots, n + l - 1$; using the periodicity of f we obtain that their versions over the indexes $\{K + 1, \ldots, T + K\}$, $(n + 1)', \ldots, (n + l - 1)'$ are such that all of them belong

to I_{δ_1}; we conclude that then $n' \in I_{\delta_1 \mathsf{U} \delta_2}$ and our algorithm would have added to G the pair (T, n'), which represents n. However, it is possible that although $m \geq n$, their version over $\{K, \ldots, T+K\}$ may verify $n' > m'$. This case is only possible because there will exist an index k between n and m such that $k - K \mod T = 0$; in this case following a similar argument as above we can conclude that $n', \ldots, k' \in I_{\delta_1}$ and $k'+1 \ldots m'-1 \in I_{\delta_1}$; using the fact that $m' \in I_{\delta_2}$ we get that $n' \in I_{\delta_1 \mathsf{U} \delta_2}$. In this case the final condition of Algorithm 1 is fulfilled and so we in fact have added to G the pair (T, n'), which represents n. □

Theorem 1. *Let δ be an EPLTL formula, and (M, μ) a Markov chain. Assume that for any subterm $t_i = 0$ of δ, there exists a representation of $Z(\mu^T M^i t_i)$ which is (F_i, G_i). Then there exists (F, G) a representation for I_δ.*

Proof. The proof follows from using structural induction. If $\delta \equiv (t_i = 0)$, then by hypothesis we already have a representation for I_δ. So, for each of the connectives, we need to show that, assuming that the connectives' subformulae have representations (and therefore are unions of finite sets and arithmetical progressions), the satisfaction set for the connective will also be the union of a finite set and arithmetical progressions. Since the proofs for all connectives are constructive, we can obtain a representation for each of them. Using Proposition 1, it remains to prove the case of the U connective. With the Algorithm 1, we obtain that we can construct a representation for any EPLTL formulae. □

Corollary 1. The verification problem of EPLTL formulae in Markov chains is equireducible to the Skolem problem.

Remark 4. The focus of this work is on decidability. While extracting the running time complexity of the procedures for each connective is straightforward, we can not make claims about the complexity class in which the verification problem lies relative to the Skolem problem.

4 Approximate Skolem Problem and verification of EPLTL formulae

As the Skolem problem is not known to be decidable, using it to verify EPLTL seems to be a doomed enterprise for the time being. However, in most applications, we are willing to accept results carrying a small error, at least to deal with finite precision representations. We intend on using the closure results for representations of EPLTL formulae for approximate model checking of Markov chains. We will need a small adjustment to the syntax and semantics of EPLTL in order to cope with error bounds.

Definition 6. For any $\epsilon \in \mathbb{Q}$, the well formed formulae in ϵ-approximate EPLTL^ϵ

over the propositional symbols Λ are given below in Backus-Naur notation:

$$\beta := p \;[\!]\; (\beta \wedge \beta) \;[\!]\; (\neg\beta) \qquad \text{(basic formulae)}$$

$$t := r \;[\!]\; (t + t) \;[\!]\; r \int \beta \qquad \text{(rational terms)}$$

$$\delta^\epsilon := \underbrace{\sim \delta^\epsilon \;[\!]\; \delta^\epsilon \cap \delta^\epsilon \;[\!]\; \mathsf{X}\delta^\epsilon \;[\!]\; \delta^\epsilon \mathsf{U} \delta^\epsilon}_{\text{temporal formulae}} \quad [\!] \quad \underbrace{(t = t)^\epsilon}_{\text{comparison formulae}} \qquad \text{(formulae)}$$

where $p \in \Lambda$ and $r \in \mathbb{Q}$.

For an EPLTL formula δ and$\epsilon \in \mathbb{Q}$, we define the syntactic translation δ^ϵ of δ in the expected way. The denotation of terms is as in Definition 3 and the satisfaction relation is as expected:

Definition 7. The satisfaction relation $\Vdash^\epsilon$ between a Markov chain $\mathcal{M}$, an instant i, and a EPLTL^ϵ formula δ^ϵ is defined inductively:

- $\mathcal{M}, i \Vdash^\epsilon (t_1 = t_2)^\epsilon$ iff $|[\![t_1]\!]_{\mathcal{M},i} - [\![t_2]\!]_{\mathcal{M},i}| < \epsilon$,
- $\mathcal{M}, i \Vdash^\epsilon (\delta_1^\epsilon \cap \delta_2^\epsilon)$ iff $\mathcal{M}, i \Vdash^\epsilon \delta_1^\epsilon$ and $\mathcal{M}, i \Vdash^\epsilon \delta_2^\epsilon$,
- $\mathcal{M}, i \Vdash^\epsilon (\sim \delta)$ iff $\mathcal{M}, i \nVdash^\epsilon \delta$,
- $\mathcal{M}, i \Vdash^\epsilon (\mathsf{X}\delta^\epsilon)$ iff $\mathcal{M}, i+1 \Vdash^\epsilon \delta^\epsilon$,
- $\mathcal{M}, i \Vdash^\epsilon (\delta_1^\epsilon \mathsf{U} \delta_2^\epsilon)$ iff there exists $j \geq i$ such that $\mathcal{M}, j \Vdash^\epsilon \delta_2^\epsilon$ and $\mathcal{M}, k \Vdash^\epsilon \delta_1^\epsilon$, for all $i \leq k < j$.

In this section our main goal is to prove the following result:

Theorem 2. *For any $\delta \in$ EPLTL and Markov chain $\mathcal{M}$, there exists a computable error margin $d(\delta, \mathcal{M})$ such that for all $\epsilon \in]0; d[\cap \mathbb{Q}$ we can decide whether $M \Vdash^\epsilon \delta^\epsilon$.*

In order to do so we will first prove the following result:

Theorem 3. *For any Markov chain $\mathcal{M} = (M, \mu)$, and any rational vector y, there exists a computable error margin d such that for all $\epsilon \in]0; d[\cap \mathbb{Q}$ we can decide whether there exists $i \in \mathbb{N}$ such that $-\epsilon < \mu^T M^i y < \epsilon$.*

Given a rational term t and a Markov chain $\mathcal{M} = (M, \mu)$, we will be interested in characterizing the set $\{i \in \mathbb{N} : -\epsilon < \mu^T M^i t < \epsilon\}$ for a suitable ϵ. In fact, this set will be the union of a finite set and a finite number of arithmetical progressions. However, unlike in the Skolem problem, we can actually compute a representation for this set (for suitable $\epsilon \in \mathbb{Q}$). Therefore, using the results already proven in last section about the temporal and propositional connectives, it will follow that the index set of δ^ϵ will also be the union of a finite set and a finite number of arithmetical progressions.

We will first show how to characterize the set $\{i \in \mathbb{N} : -\epsilon < \mu^T M^i t < \epsilon\}$, for a suitable precision ϵ. Note first that, using Jordan decomposition:

$$\begin{aligned}\mu^T M^n t &= \mu^T (SDS^{-1})^n t^* \\ &= \mu^T S D^n S^{-1} t^* \\ &= \sum_j p_j(n) \lambda_j^n\end{aligned}$$

The polynomials $p_j(n)$ have degree bounded by the size of the matrix. Notice that these polynomials will, in general, be complex and that the eigenvalues may also be complex; however one needs to remember that, in the end, the sum must still be a rational value.

Perron-Frobënius Theorem applied to irreducible stochastic matrices allows us to state the following:

- All eigenvalues verify $|\lambda_j| \leq 1$;
- There exists at least one eigenvalue j such that $\lambda_j = 1$;
- Other eigenvalues of absolute value 1 are all the roots of 1 for some degree.

In general, our stochastic matrix M is not necessarily irreducible; however, if we rewrite it using permutation matrices, we can obtain a matrix in upper-triangular block form such that:

- $PMP^{-1} = \begin{pmatrix} B_1 & ... & ... & ... \\ \mathbf{0} & B_2 & ... & ... \\ \mathbf{0} & \mathbf{0} & ... & ... \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & B_n \end{pmatrix}$
- Each matrix B_i is either stochastic and irreducible, or substochastic and irreducible.

The spectrum of M is the union of the spectra of each component B_i; therefore applying Perron Frobënius for each B_i, we conclude that the spectrum of M can be divided in several sets of all roots of unity for some degrees (which have absolute value 1), and all other eigenvalues (which have absolute value less than 1).

Then, we can expand the summation as follows:

$$
\begin{aligned}
\mu^T M^n t &= \sum_j p_j(n)\lambda_j^n \\
&= \sum_{\{j:|\lambda_j|=1\}} p_j(n)\lambda_j^n + \sum_{\{j:|\lambda_j|<1\}} p_j(n)\lambda_j^n \\
&= \sum_{\{j:|\lambda_j|=1\}} p_j(n) r_j^n e^{i\theta_j n} + \sum_{\{j:|\lambda_j|<1\}} p_j(n) r_j^n e^{i\theta_j n} \\
&= \underbrace{\sum_{\{j:|\lambda_j|=1\}} p_j(n) e^{2\pi i \frac{p_j}{q_j} n}}_{P(n)} + \underbrace{\sum_{\{j:|\lambda_j|<1\}} p_j(n) r_j^n e^{i\theta_j n}}_{D(n)}.
\end{aligned}
$$

Lemma 4. Let $\mathcal{M} = (M, \mu)$ be a Markov chain, t the vector associated with EPLTL term t (as in Remark 3). Then $P(n)$, and $D(n)$, defined as above are such that:

- $P(n)$ is periodic,
- $|D(n)| \leq q(n)R^n$, for some $0 \leq R < 1$ and polynomial $q(n)$, such that $q(n)R^n$ is monotonically decreasing after some $m \in \mathbb{N}$.
- $P(n), D(n) \in \mathbb{R}$.

Proof. We start by showing that $|D(n)| \leq q(n)R^n$, for some $0 \leq R < 1$ and polynomial $q(n)$:

$$
\begin{aligned}
|D(n)| &= |\sum_{\{j:|\lambda_j|<1\}} p_j(n) r_j^n e^{i\theta_j n}| \\
&\leq \sum_{\{j:|\lambda_j|<1\}} |p_j(n)| r_j^n \\
&\leq q(n)R^n.
\end{aligned}
$$

With, for instance $q(n) = c(n^d + 1)$, for $d = \max \deg(p_j(n))$ and sufficiently large $c > 0$, and $R \in]0;1[\cap \mathbb{Q}$ chosen s.t. $r_j < R$; therefore not only we can compute m s.t. $q(n)R^N$ is monotonically decreasing but also $\lim |D(n)| = 0$.

In order to prove the other two assertions, let T be the least common multiple of all the denominators of the roots of unity in $P(n)$. Furthermore, let $P_s(k)$ and $D_s(k)$ with $s \in \{0, 1, \ldots, T-1\}$ be defined as:

$$
P_s(k) = P(s + Tk), D_s(k) = D(s + Tk), k \in \mathbb{N}.
$$

The first thing to note is that for each $s \in \{0, 1, \ldots, T-1\}$, $P_s(k)$ is a

complex polynomial on k:

$$\begin{aligned}
P_s(k) &= \sum_{\{j:|\lambda_j|=1\}} p_j(s+Tk)e^{2\pi i\frac{p_j}{q_j}(s+Tk)}, \\
&= \sum_{\{j:|\lambda_j|=1\}} p_j(s+Tk)e^{2\pi i\frac{p_j}{q_j}s}e^{2\pi i\frac{p_j}{q_j}Tk}, \\
&= \sum_{\{j:|\lambda_j|=1\}} p_j(s+Tk)e^{2\pi i\frac{p_j}{q_j}s}, \\
&= \sum_{\{j:|\lambda_j|=1\}} p_j(s+Tk)e^{2\pi i\frac{p_j}{q_j}s}.
\end{aligned}$$

Note that $p_j(s+Tk)$ are complex polynomials on k, $e^{2\pi i\frac{p_j}{q_j}s}$ are complex constants, and therefore $P_s(k)$ is a complex polynomial on k, too. So, in fact $P_s(k) = \sum a_j k^j + i\sum b_j k^j$ for some real coefficients a_j, b_j. Now, for $k \in \mathbb{N}$, the imaginary part of $P_s(k)$, $\sum b_j k^j$, is a real polynomial on k that is either constant or tends towards $\pm\infty$. However, we know that $\lim D_s(k) = 0$ and so both $\lim \mathrm{Re}(D_s(k)) = 0$ and $\lim \mathrm{Im}(D_s(k)) = 0$. Therefore, if $\lim \mathrm{Im}(P_s(k)) = \pm\infty$, then $\mu^T M^{s+Tk}t = \mathrm{Re}(P_s(k)) + i\mathrm{Im}(P_s(k)) + D_s(k)$ would have to be complex, which is impossible.

Suppose now that $\mathrm{Im}(P_s(k)) = c$, with $c \neq 0$. The same reasoning follows since $\mathrm{Im}(D_s(k))$ would need to be $-c$ in order for $\mu^T M^{s+Tk}t$ to be a rational value, which clearly is impossible because $\lim \mathrm{Im}(D_s(k)) = 0$. Therefore $P_s(k)$ is a real polynomial on k. This also entails that $D_s(k)$ has to be real (algebraic) valued.

We will now prove that $P_s(k)$ is in fact constant. So far we have shown that $P_s(k)$ is a real sequence. Suppose $P_s(k)$ is not a constant; then $\lim |P_s(k)| = \infty$. Naturally, this implies that $\lim |P_s(k) + D_s(k)| = \infty$, since $\lim |D_s(k)| = 0$. Note, however, that $|\mu^T M^n t| = |(\mu^t M^n).t)| \leq L||t||_1$, with L the size of the matrix, because $(\mu^T M^n)$ is a probabilistic vector. This implies that $P_s(k)$ is bounded and therefore, being a polynomial, has to be constant;

This argument holds for each s and as such, the whole sequence $P(n)$ has to be periodic. □

We now show that we can characterize the set $\{n \in \mathbb{N} : -\epsilon < \mu^T M^n t < \epsilon\}$:

Theorem 4. *There exists d, dependent on the Markov chain $\mathcal{M} = (M, \mu)$, and term t, such that, for all $\epsilon \in]0, d[$ we can compute a representation (F, G) for $\{n \in \mathbb{N} : -\epsilon < \mu^T M^n t < \epsilon\}$.*

Proof. Consider $P(n)$ and $D(n)$ as computed above. We are interested in characterizing the set $\{n \in \mathbb{N} : -\epsilon < P(n) + D(n) < \epsilon\}$. Recalling that $P(n)$ is periodic with a certain period T (computed in Lemma 4), we will consider $P_s(k) = P(s+Tk)$ and $D_s(k) = D(s+Tk)$, for $s \in \{0, \ldots, T-1\}$. We will also compute $q(n)$ and R to be such that $|D(n)| < q(n)R^n$ (as in Lemma 4). Moreover, we know that $q(n)R^n$ is decreasing after some index m.

Set $d = \min_{s\in\{0,\ldots,T-1\}}\{|P_s(0)| : P_s(0) \neq 0\}$; for any $\epsilon \in]0; d[$ and for any residue class s, there are two possibilities: either $P_s(0) = 0$, (in which case $P_s(0) = P_s(k) \in]-\epsilon, \epsilon[$, for all k), or $P_s(0) = P_s(k) \notin]-\epsilon, \epsilon[$ for any k. Now, we will collect the indexes such that $\{k \in \mathbb{N} : -\epsilon < P_s(k) + D_s(k) < \epsilon\}$ for each residue class s.

Suppose that the residue class of s does not lie within $]-\epsilon, \epsilon[$. We note that then for almost all k, $P_s(k) + D_s(k) \notin]-\epsilon, \epsilon[$, since $\lim P_s(k) + D_s(k) = P_s(0)$; so, it is a simple matter of collecting all the indexes in this class of residues that fall into $]-\epsilon, \epsilon[$ until m or $q(s+Tk)R^{s+Tk} < |P_s(0) - \epsilon|$, whichever comes later. In this case, the indexes so collected are added to the finite part of the representation F, and we are guaranteed that no more exceptional zeroes my occur in this residue class.

Suppose that the residue class of s lies within $]-\epsilon, \epsilon[$, or equivalently $P_s(0) = 0$. Then, we know that, after m, when $q(s + Tk)R^{s+Tk} < \epsilon$, then all the subsequent indexes ($i > k$) will lie inside $]-\epsilon, \epsilon[$. In this case, we compute this index k and add the arithmetical progression $(T, Tk + s)$ to the arithmetical part of the representation, G. Furthermore, from 0 to k we collect the indexes i that still lie inside $]-\epsilon, \epsilon[$ and add $Ti+s$ to the finite part of the representation, F. □

Remark 5 (**Characterizing the set** $\{n \in \mathbb{N} : -\epsilon < P(n) + D(n)\}$). While our syntax of EPLTL does not allow inequalities, we could extend it so that $(t_1 < t_2)^\epsilon$ would semantically mean $[\![t_1 - t_2]\!] < \epsilon$, and $(t_1 > t_2)^\epsilon$ would semantically mean $[\![t_1 - t_2]\!] > -\epsilon$. The proof of Theorem 4 can be adapted for these cases. Without loss of generality we sketch the case for $t > -\epsilon$: we consider the same d; if $P_s(0)$ falls into the interval $]-\epsilon; +\infty[$ we know that almost always $P_s(k) + D_s(k)$ will belong to that interval. Therefore we would need to collect the exceptional indexes that may appear up to the point after m, where $q(s + Tk)R^{s+Tk} < |P_s(0) - \epsilon|$, adding the arithmetical progression afterwards. Otherwise, if $P_s(0)$ falls in $]-\infty, -\epsilon]$, we note that we are guaranteed that $P_s(0) \neq -\epsilon$ and we just need to compute $P_s(k) + D_s(k)$ up to the point where we are certain that the remaining sequence will be in $]-\infty, -\epsilon]$, which which will happen when $q(s + Tk)R^{s+Tk} < \epsilon$.

With this result, we can now use the results about the characterizations of satisfaction sets from preceding section to verify EPLTL^ϵ formulae. One must take into account that in the case of multiples comparison terms $t_i = 0$, the choice of ϵ must be smaller than the minimum of all d computed in Theorem 4.

Theorem 5. *Given a Markov chain (M, μ), and formula $\delta \in$ EPLTL, there exists $d > 0$, such that for all $\epsilon \in]0, d[\cap\mathbb{Q}$ we can decide $(M, \mu) \Vdash^\epsilon \delta^\epsilon$.*

Remark 6. The focus of this work is on decidability. However, we make some comments on the computational complexity of the procedure presented for the approximate Skolem problem. The algorithm may be divided in two parts: (i) finding $d = \min_{s\in\{0,\ldots,T-1\}}\{|P_s(0)| : P_s(0) \neq 0\}$ from M, μ and t and, (ii) given ϵ, collecting the indexes such that $\{k \in \mathbb{N} : -\epsilon < P(n) + D(n) < \epsilon\}$. In order

to compute d, we need obtain a Jordan decomposition of M (which allows us to rewrite the problem in terms of $P(n) + D(n)$), find the eigenvalues that are roots of unity, compute the least common multiple of their denominators, T and, finally, evaluate $P(n)$ from 0 to $T-1$, obtaining $P_s(0)$. All these operations are of well known complexity [6].

The analysis of the relation between d and ϵ is more interesting. For each $s \in \{0, \ldots, T-1\}$, we need to explicitly compute $P_s(k) + D_s(k)$ up to a certain index to account for the possibility of exceptional zeroes. For each residue class $s \in \{0, \ldots, T-1\}$ we either have to compute $P_s(k) + D_s(k)$ up to $q(s + Tk)R^{s+Tk} < |P_s(0) - \epsilon|$ or $q(s+Tk)R^{s+Tk} < \epsilon$. For the sake of simplifying the analysis, let us suppose that the procedure, for each s, naively checks all the indexes until both $q(s + Tk)R^{s+Tk} < |d - \epsilon|$ and $q(s + Tk)R^{s+Tk} < \epsilon$. This clearly requires at least as many checks as the proposed method.

Now the question becomes how many checks do we have to make until $q(k)R^k < \min\{|d - \epsilon|, \epsilon\}$. It is clear that the procedure becomes harder when $m = \frac{1}{\min\{|d-\epsilon|,\epsilon\}}$ increases. Let $N(m) = \min_{n\in\mathbb{N}} q(n)R^n < \frac{1}{m} = \min\{|d - \epsilon|, \epsilon\}$ be the number of exceptional indexes that we need to check. We will show that $N(m) \in O(\log m)$ by establishing that $\limsup_{m\to\infty} N(m) < k \log m$ for some k. Consider the function $f(m) = q(k \log m)R^{k \log m}$. We will show, that for some k, $\lim f(m) < \frac{1}{m}$. Since, for all m, $N(m)$ is the first value such that $q(n)R^n \leq \frac{1}{m}$, after some m, $N(m) \leq k \log m$.

So, it remains to show that $\lim f(m) < \frac{1}{m}$. $f(m) = q(k \log m)R^{k \log m} = q(k \log m)\frac{1}{m^{\log \frac{1}{R^k}}}$. Since the degree of q is fixed and $R < 1$, we can choose k large enough to guarantee that $\frac{1}{m^{\log \frac{1}{R^k}}}$ decreases faster than $mq(k \log m)$. Therefore, $N(m) \in O(\log m)$, that is, the number of exceptional indexes that have to be checked varies logarithmically with the inverse of the distance to the bounds of the interval.

5 Conclusions and Future Work

In this work we have shown the connection between the Skolem problem and the verification problem for temporal properties on probabilistic dynamical systems induced by Markov chains. Since significant advances on the decidability (or lack thereof) of the Skolem problem seem unlikely in the near future [19], we turned our attention towards approximate versions of these problems. In this context, we have presented procedures to decide the problems. While the focus of this work was on decidability, a natural question is to wonder about the complexity of the approximate problems. We have made some comments on the running time of the presented algorithms but have not investigated lower complexity bounds for the decision problems or the reduction of the verification problem to the Skolem problem.

Another natural progression for this work would be an implementation of the verification algorithm for the approximate version of the problem. This approach ties to the complexity analysis since, in this work, we were more

concerned with clarity of exposition of the procedures rather than efficiency.

References

[1] M. Agrawal, S. Akshay, B. Genest, and P. S. Thiagarajan. Approximate verification of the symbolic dynamics of Markov chains. In *LICS*, pages 55–64. IEEE, 2012.

[2] M. Agrawal, S. Akshay, B. Genest, and P.S. Thiagarajan. Approximate verification of the symbolic dynamics of Markov chains. *Journal of the ACM (JACM)*, January 2014.

[3] C. Baier, E. M. Clarke, V. Hartonas-Garmhausen, M. Z. Kwiatkowska, and M. Ryan. Symbolic model checking for probabilistic processes. In *ICALP*, volume 1256 of *LNCS*, pages 430–440. Springer, 1997.

[4] C. Baier and J.-P. Katoen. *Principles of model checking*. MIT Press, 2008.

[5] P. Baltazar and P. Mateus. Temporalization of probabilistic propositional logic. In S. Artemov and A. Nerode, editors, *Logic Foundations of Computer Science 2009*, volume 5407 of *Lecture Notes in Computer Science*, pages 46–60. Springer, 2009.

[6] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[7] D. Beauquier, A. Rabinovich, and A. Slissenko. A logic of probability with decidable model checking. *Journal of Logic and Computation*, 16(4):461–487, 2006.

[8] J. Berstel and M. Mignotte. Deux propriétés décidables des suites récurrentes linéaires. *Bulletin de la Société Mathématique de France*, 104:175–184, 1976.

[9] F. Ciesinski and C. Baier. Liquor: A tool for qualitative and quantitative linear time analysis of reactive systems. In *QEST'06*, pages 131–132. IEEE CS Press, 2006.

[10] F. Ciesinski and M. Größer. On probabilistic computation tree logic. In *Validation of Stochastic Systems*, volume 2925 of *LNCS*, pages 147–188. Springer, 2004.

[11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, second edition edition, 2001.

[12] J.-P. Katoen, E. M. Hahn, H. Hermanns, D. Jansen, and I. Zapreev. The ins and outs of the probabilistic model checker MRMC. In *QEST'09*, pages 167–176. IEEE CS Press, 2009.

[13] V. Korthikanti, M. Viswanathan, G. Agha, and Y. Kwon. Reasoning about MDPs as transformers of probability distributions. In *QEST'10*, pages 199–208. IEEE CS Press, 2010.

[14] M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In *CAV'11*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011.

[15] Y. Kwon and G. Agha. Linear inequality ltl (iltl): A model checker for discrete time Markov chains. In *ICFEM*, volume 3308 of *LNCS*, pages 194–208. Springer, 2004.

[16] K. Mahler and J. W. S. Cassels. On the Taylor coefficients of rational functions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52:39–48, 1956.

[17] P. Mateus and A. Sernadas. Weakly complete axiomatization of exogenous quantum propositional logic. *Information and Computation*, 204(5):771–794, 2006. ArXiv math.LO/0503453.

[18] J. Ouaknine and J. Worrell. Decision problems for linear recurrence sequences. In *RP*, volume 7550 of *LNCS*, pages 21–28. Springer, 2012.

[19] J. Ouaknine and J. Worrell. Positivity problems for low-order linear recurrence sequences. *Computing Research Repository*, abs/1307.2779, 2013.

[20] T. Skolem. Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen und diophantischer Gleichungen. 8. Skand. Mat. Kongr., Stockhohn, 1934, 163-188 (1934)., 1934.

[21] M. Y. Vardi. Probabilistic linear-time model checking: An overview of the automata-theoretic approach. In *ARTS*, volume 1601 of *LNCS*, pages 265–276. Springer, 1999.