On Decidability of Time-Bounded Reachability in CTMDPs

Rupak Majumdar

Max-Planck Institute for Software Systems, Kaiserslautern, Germany
rupak@mpi-sws.org

Mahmoud Salamati

Max-Planck Institute for Software Systems, Kaiserslautern, Germany
msalamati@mpi-sws.org

Sadegh Soudjani

Newcastle University, Newcastle upon Tyne, UK
sadegh.soudjani@newcastle.ac.uk

Abstract

We consider the time-bounded reachability problem for continuous-time Markov decision processes. We show that the problem is decidable subject to Schanuel's conjecture. Our decision procedure relies on the structure of optimal policies and the conditional decidability (under Schanuel's conjecture) of the theory of reals extended with exponential and trigonometric functions over bounded domains. We further show that any unconditional decidability result would imply unconditional decidability of the bounded continuous Skolem problem, or equivalently, the problem of checking if an exponential polynomial has a non-tangential zero in a bounded interval. We note that the latter problems are also decidable subject to Schanuel's conjecture but finding unconditional decision procedures remain longstanding open problems.

2012 ACM Subject Classification Theory of computation → Numeric approximation algorithms; Mathematics of computing → Markov processes; Theory of computation → Verification by model checking

Keywords and phrases CTMDP, Time bounded reachability, Continuous Skolem Problem, Schanuel's Conjecture

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.133

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Funding This research was funded in part by the Deutsche Forschungsgemeinschaft project 389792660-TRR 248 and by the European Research Council under the Grant Agreement 610150 (ERC Synergy Grant ImPACT).

Acknowledgements We thank Joël Ouaknine, James Worrell, and Joost-Pieter Katoen for discussions and pointers.

1 Introduction

Continuous-time Markov decision processes (CTMDPs) are a widely used model for continuous-time systems which exhibit both stochastic and non-deterministic choice. A CTMDP consists of a finite set of states, a finite set of actions, and for each action, a transition rate matrix that determines the rate (in an exponential distribution in continuous time) to go from one state to the next when the action is chosen. A *policy* for a CTMDP maps a timed execution path to state-dependent actions. Given a fixed policy, a CTMDP determines a stochastic process in continuous time, where the rate matrix determines the distribution of switches.

A fundamental decision problem for CTMDPs is the *time-bounded reachability problem*, which asks, given a CTMDP $\mathcal{M}$ with a designated "**good**" state, a time bound B, and a rational vector r, whether there exists a policy that controls the Markov decision process such that the probability of reaching the good state from state s within time bound B is at least $r(s)$. The time-bounded reachability problem is at the core of model checking CTMDPs with respect to stochastic temporal logics [5] and has been extensively studied [10, 21, 28, 20, 9].

Existing papers either consider *time-abstract* policies [5, 25, 8, 28, 20] or focus on numerical approximation schemes [10, 21, 3, 13, 9, 26]. However, policies that depend on time are strictly more powerful and the *decision problem* has remained open. For the special case of continuous-time Markov chains (CTMCs), where each state has a unique action, the time-bounded reachability problem is decidable [4]. The proof uses tools from transcendental number theory, specifically, the Lindemann-Weierstrass theorem. One might expect that a similar argument might be used to show decidability for CTMDPs as well.

In this paper, we show *conditional* decidability. Our result uses, like several other conditional results on dynamical systems, Schanuel's conjecture from transcendental number theory (see, e.g., [14]). Our proof has the following ingredients. First, we use the fact that the optimal policy for the time-bounded reachability problem is a timed, *piecewise constant* function with a *finite* number of switches [19, 22, 24]. We show that each switch point of an optimal policy corresponds to a non-tangential zero of an associated linear dynamical system. Second, we use the result of Macintyre and Wilkie [16, 17] that Schanuel's conjecture implies the decidability of the real-closed field together with the exponential, sine, and cosine functions over a bounded domain. The existence of non-tangential zeros of linear dynamical systems can be encoded in this theory. Third, for each natural number $k \in \mathbb{N}$, we write a sentence in this theory whose validity implies there is an optimal strategy with exactly k switch points. By enumerating over k, we find the exact number of switches in an optimal strategy. Finally, we write another sentence in the theory that checks if the reachability probability attained by (an encoding of) the optimal policy is greater than the given bound.

We also study the related decision problem whether there is a *stationary* (i.e., time independent) optimal policy. We show that there is a "direct" conditional decision procedure for this problem based on Schanuel's conjecture and recent results on zeros of exponential polynomials [11], which avoids the result of Macintyre and Wilkie.

At the same time, we show that an *unconditional* decidability result is likely to be very difficult. We show that the bounded continuous-time Skolem problem [7, 11] reduces to checking if there is an optimal stationary policy in the time-bounded CTMDP problem. The bounded continuous Skolem problem is a long-standing open problem about linear dynamical systems [11, 7]; it asks if a linear dynamical system in continuous time has a non-tangential zero in a bounded interval. Our reduction, in essence, demonstrates that CTMDPs can "simulate" any linear dynamical system: a non-tangential zero in the dynamics corresponds to a policy switch point in the simulating CTMDP.

Our result is in the same spirit as several recent results providing conditional decision procedures, based on Schanuel's conjecture, or hardness results, based on variants of the Skolem problem, for problems on probabilistic systems. For example, Daviaud et al. [12] showed conditional decidability of subcases of the containment problem for probabilistic automata subject to the conditional decidability of the theory of real closed fields with the exponential function [18, 27]. For lower bounds, Akshay et al. [2] showed a reduction from the (unbounded, discrete) Skolem problem to reachability on discrete time Markov chains and Piribauer and Baier [23] show that the positivity problem in discrete time can be reduced into several decision problems corresponding to optimization tasks over discrete time MDPs.

In summary, we summarize our contribution as the following theorem.

▶ **Theorem 1.** *(1) The time-bounded reachability problem for CTMDPs is decidable assuming Schanuel's conjecture. (2) Whether the time-bounded reachability problem has a stationary optimal policy is decidable assuming Schanuel's conjecture. (3) The bounded continuous Skolem problem reduces to checking if the time-bounded reachability problem has a stationary optimal policy.*

2 Continuous Time Markov Decision Processes

▶ **Definition 2.** *A continuous-time Markov decision process (CTMDP) is a tuple $\mathcal{M} = (S, \mathcal{D}, \mathbf{Q})$ where*

- $S = \{1, 2, \dots, \mathfrak{n}\}$ *is a finite set of states for some* $\mathfrak{n} > 0$*;*
- *a set* $\mathcal{D} = \prod_{s=1}^{\mathfrak{n}} \mathcal{D}_s$ *of* decision vectors*, where* $\mathcal{D}_s$ *is a finite set of* actions *that can be taken in state* $s \in S$*;*
- $\mathbf{Q}$ *is a* $\mathcal{D}$*-indexed family of* $\mathfrak{n} \times \mathfrak{n}$ generator matrices*; we write* $\mathbf{Q}^{\mathbf{d}}$ *for the generator matrix corresponding to the decision vector* $\mathbf{d} \in \mathcal{D}$*. The entry* $\mathbf{Q}^{\mathbf{d}}(s, s') \geq 0$ *for* $s' \neq s$ *gives the* rate of transition *from state* s *to state* s' *under action* $\mathbf{d}(s)$*, and* $\mathbf{Q}^{\mathbf{d}}(s, s')$ *is independent of elements of* $\mathbf{d}$ *except* $\mathbf{d}(s)$*. The entry* $\mathbf{Q}^{\mathbf{d}}(s, s) = -\sum_{s' \neq s} \mathbf{Q}^{\mathbf{d}}(s, s')$*.*

A CTMDP $\mathcal{M} = (S, \mathcal{D}, \mathbf{Q})$ with $|\mathcal{D}| = 1$, i.e., when only a unique action can be taken in each state, is called a *continuous-time Markov chain* (CTMC) and is simply denoted by the tuple $(S, \mathbf{Q})$, and with abuse of notation, we also write $\mathbf{Q}$ for the unique generator matrix. The CTMDP $\mathcal{M}$ reduces to a CTMC whenever a decision vector $\mathbf{d}$ is fixed for all time on the CTMDP.

Intuitively, $\mathbf{Q}^{\mathbf{d}}(s, s') > 0$ indicates that by fixing a decision vector $\mathbf{d}$, a transition from s to s' is possible and that the timing of the transition is exponentially distributed with rate $\mathbf{Q}^{\mathbf{d}}(s, s')$. If there are several states s' such that $\mathbf{Q}^{\mathbf{d}}(s, s') > 0$, more than one transition will be possible. For each decision vector $\mathbf{d} \in \mathcal{D}$ and any $s \in S$, the total rate of taking an outgoing transition from state s when $\mathbf{d}$ is fixed is given by $E_{\mathbf{d}}(s) = \sum_{s' \neq s} \mathbf{Q}^{\mathbf{d}}(s, s')$, By fixing this decision vector $\mathbf{d}$, a transition from a state s into s' occurs within time t with probability

$$\mathbf{P}(s, s', t) = \frac{\mathbf{Q}^{\mathbf{d}}(s, s')}{E_{\mathbf{d}}(s)}.(1 - e^{-E_{\mathbf{d}}(s)t}), \quad t \geq 0.$$

Intuitively, $1 - e^{-E_{\mathbf{d}}(s)t}$ is the probability of taking an outgoing transition at s within time t (exponentially distributed with rate $E_{\mathbf{d}}(s)$) and $\mathbf{Q}^{\mathbf{d}}(s, s')/E_{\mathbf{d}}(s)$ is the probability of taking transition to s' among possible next states at s. Thus, the total probability of moving from s to s' under the decision $\mathbf{d}$ in one transition, written $\mathbf{P}_{\mathbf{d}}(s, s')$ is $\mathbf{Q}^{\mathbf{d}}(s, s')/E_{\mathbf{d}}(s)$. A state $s \in S$ is called *absorbing* if and only if $\mathbf{Q}^{\mathbf{d}}(s, s') = 0$ for all $s' \in S$ and all decision vectors $\mathbf{d} \in \mathcal{D}$. For an absorbing state, we have $E_{\mathbf{d}}(s) = 0$ for any decision vector $\mathbf{d}$ and no transitions are enabled. The initial state of a CTMDP is either fixed deterministically or selected randomly according to a probability distribution α over the set of states S.

Consider a time interval $[0, B]$ with time bound $B > 0$. Let Ω denote the set of all right-continuous step functions $f : [0, B] \to S$, i.e., there are time points $t_0 = 0 < t_1 < t_2 < \ldots < t_m = B$ such that $f(t') = f(t'')$ for all $t', t'' \in [t_i, t_{i+1})$ for all $i \in \{0, 1, \dots, m-1\}$. Let $\mathcal{F}$ denote the sigma-algebra of the *cylinder sets*

$$\mathsf{Cyl}(s_0, I_0, \dots, I_{m-1}, s_m) := \{f \in \Omega \,|\, \forall 0 \leq i \leq m \cdot f(t_i) = s_i \wedge i < m \Rightarrow (t_{i+1} - t_i) \in I_i\}. \quad (1)$$

for all m, $s_i \in S$ and non-empty time intervals $I_0, I_1, \dots, I_{m-1} \subset [0, B]$.

▶ **Definition 3.** *A* policy π *is a function from* $[0, B]$ *into* $\mathcal{D}$*, which is assumed to be Lebesgue measurable. Any policy gives a decision vector* $\pi_t \in \mathcal{D}$ *at time* t *such that the action* $\pi_t(s)$ *is taken when the CTMDP is at state* s *at time* t*. The set of all such polices is denoted by* Π_B*.*

Any policy π together with an initial distribution α induces the probability space $(\Omega, \mathcal{F}, \mathbf{P}^\pi_\alpha)$. If the initial distribution is chosen deterministically as $s \in S$, we denote the probability measure by $\mathbf{P}^\pi_s$ instead of $\mathbf{P}^\pi_\alpha$.

A policy $\pi : [0, B] \to \mathcal{D}$ is *piecewise constant* if there exist a number $m \in \mathbb{N}$ and time points $t_0 = 0 < t_1 < t_2 < \ldots < t_m = B$ such that $\pi_{t'} = \pi_{t''}$ for all $t', t'' \in (t_i, t_{i+1}]$ and all $i \in \{0, 1, \ldots, m-1\}$. The policy is *stationary* if $m = 1$. We denote the class of stationary policies by $\Pi_{\mathfrak{st}}$; observe that a stationary policy is given by a fixed decision vector, so $\Pi_{\mathfrak{st}}$ is isomorphic with the set of decision vectors $\mathcal{D}$. In particular, it is a finite set.

▶ Remark 4. The policies in Def. 3 are called *timed positional* policies since the action is selected deterministically as a function of time and the state of the CTMDP at that time. A stationary policy is only positional since the selected action is independent of time.

▶ **Problem 1.** *Consider a CTMDP* $\mathcal{M} = (\{1, \ldots, n\} \uplus \{\mathbf{good}\}, \mathcal{D}, \mathbf{Q})$ *with a distinguished absorbing state named* **good** *and a time bound* $B > 0$*. Define the event*

$$\boldsymbol{reach} := \cup\{f \in \Omega \mid f(t) = \mathbf{good} \text{ for some } t \in [0, B]\}. \tag{2}$$

The time-bounded reachability problem *asks if for a rational vector* $r \in [0, 1]^n$*, we have*

$$\sup_{\pi \in \Pi_B} \mathbf{P}^\pi_s(\boldsymbol{reach}) > r(s), \qquad \text{for all } s \in \{1, \ldots, n\}.$$

The event **reach** defined in (2) is written as a union of an uncountable number of functions but it is measurable in the probability space $(\Omega, \mathcal{F}, \mathbf{P}^\pi_\alpha)$ for any α. Since the state space is finite, **reach** can be written as a countable union of cylinder sets in the form of (1) by taking all the time intervals to be $[0, B]$ and enumerating over all possible sequence of states (which is countable) [6].

A policy $\pi^* \in \Pi_B$ is *optimal* if $P^{\pi^*}_s(\mathbf{reach}) = \sup_{\pi \in \Pi_B} \mathbf{P}^\pi_s(\mathbf{reach})$. Note that there are more general classes of policies that may depend also on the history of the states in the previous time points and which map the history to a distribution over $\mathcal{D}$. It is shown that piecewise constant timed positional policies are sufficient for the optimal reachability probability [19, 22, 24]. That is, if there is an optimal policy from the larger class of policies, there is already one from the class of piecewise constant, timed, positional policies.

A closely related problem is the existence of *stationary* optimal policies; here, it is possible that the optimal stationary policy performs strictly worse than an optimal policy.

▶ **Problem 2.** *Consider a CTMDP* $\mathcal{M} = (\{1, \ldots, n\} \uplus \{\mathbf{good}\}, \mathcal{D}, \mathbf{Q})$ *and a time bound* $B > 0$*. Decide whether there is an optimal policy* π^* *that is stationary, namely*

$$\exists \pi^* \in \Pi_{\mathfrak{st}} \text{ s.t. } \sup_{\pi \in \Pi_B} \mathbf{P}^\pi_s(\boldsymbol{reach}) = \mathbf{P}^{\pi^*}_s(\boldsymbol{reach}), \qquad \text{for all } s \in \{1, \ldots, n\}.$$

In the following, we shall assume that the CTMDPs and all bounds in the above decision problems are given using rational numbers. That is, rates of transitions in each generator matrix is a rational number, and the time bound B is a rational number.

▶ **Theorem 5** ([10, 19])**.** *A policy* $\pi \in \Pi_B$ *is optimal if* $\mathbf{d}_t$*, the decision vector taken by* π *at time* $B - t$*, maximizes for almost all* $t \in [0, B]$

$$\max_{\mathbf{d}_t}(\mathbf{Q}^{\mathbf{d}_t} W^\pi_t) \text{ with } \frac{d}{dt} W^\pi_t = \mathbf{Q}^{\mathbf{d}_t} W^\pi_t, \tag{3}$$

with the initial condition $W^\pi_0(\mathbf{good}) = 1$ *and* $W^\pi_0(s) = 0$ *for all* $s \in \{1, 2, \ldots, n\}$*. There exists a piecewise constant policy* π *that maximizes the equations.*

The maximization in Equation (3) above is performed element-wise. Equation (3) should be solved forward in time to construct the policy π backward in time due to the definition $\mathbf{d}_t = \pi_{B-t}$. One can alternatively write down (3) directly backward in time based on π_t.

The proof of Theorem 5 is constructive [10, 19] and is based on the following sets for any vector W:

$$\begin{aligned}
&\mathcal{F}_1(W) = \{\mathbf{d} \in \mathcal{D} \,|\, \mathbf{d} \text{ maximizes } \mathbf{Q}^{\mathbf{d}} W\},\\
&\mathcal{F}_2(W) = \{\mathbf{d} \in \mathcal{F}_1(W) \,|\, \mathbf{d} \text{ maximizes } [\mathbf{Q}^{\mathbf{d}}]^2 W\},\\
&\dots\\
&\mathcal{F}_j(W) = \{\mathbf{d} \in \mathcal{F}_{j-1}(W) \,|\, \mathbf{d} \text{ maximizes } [\mathbf{Q}^{\mathbf{d}}]^j W\}.
\end{aligned} \tag{4}$$

The sets $\mathcal{F}_j(W)$ form a sequence of decreasing sets such that $\mathcal{F}_1(W) \supseteq \mathcal{F}_2(W) \supseteq \ldots \supseteq \mathcal{F}_{n+2}(W) = \mathcal{F}_{n+k}(W)$ for all $k > 2$. An optimal piecewise constant policy is the one that satisfies the condition $\mathbf{d}_t \in \mathcal{F}_{n+2}(W_t^\pi)$ for all $t \in [0, B]$. Note that if $\mathcal{F}_j(W_t^\pi)$ has only one element for some j, $\mathcal{F}_k(W_t^\pi) = \mathcal{F}_j(W_t^\pi)$ for all $k \geq j$ and that element is the optimal decision vector. The next proposition shows that when $\mathcal{F}_{n+2}(W_t^\pi)$ has more than one element, we can pick any one (and in fact, switch between them arbitrarily).

▶ **Proposition 6.** *Let π be an optimal policy satisfying Equation (3). Take any t^* such that $\mathcal{F}_{n+2}(W_{t^*}^\pi) \neq \lim_{t\uparrow t^*} \mathcal{F}_{n+2}(W_t^\pi)$. If $\mathcal{F}_{n+2}(W_{t^*}^\pi) = \{\mathbf{d}^1, \mathbf{d}^2, \ldots, \mathbf{d}^p\}$ for some $p > 1$ and*

$$\Delta_i := \sup\{\delta > 0 \,|\, \mathbf{d}^i \in \mathcal{F}_{n+2}(W_t^\pi) \text{ for all } t \in [t^*, t^* + \delta)\}, \quad \forall i \in \{1, 2, \ldots, p\}$$

Then, $\Delta_1 = \Delta_2 = \cdots = \Delta_p$.
Suppose there are points δ_1, δ_2 such that $t^* \leq \delta_1 < \delta_2 < t^* + \Delta_1$ and for all $t \in [\delta_1, \delta_2)$, we have $\pi_{B-t} = \mathbf{d}$ for some $\mathbf{d} \in \mathcal{F}_{n+1}(W_{t^*}^\pi)$. If π' is a policy that agrees with π on $[0, \delta_1)$ but for all $t \in [\delta_1, \delta_2)$, we have $\pi'_{B-t} = \mathbf{d}'$ for some $\mathbf{d}' \in \mathcal{F}_{n+1}(W_{t^*}^\pi) \setminus \{\mathbf{d}\}$, then π' also satisfies Equation (3) for almost all $t \in [0, \delta_2)$.

Proof. Since $\mathcal{F}_{n+2}(W_{t^*}^\pi) = \mathcal{F}_{n+k}(W_{t^*}^\pi)$ for all $k > 2$, for any $\mathbf{d}^i$ and $\mathbf{d}^j$ belonging to the set $\mathcal{F}_{n+2}(W_{t^*}^\pi)$, we have $[\mathbf{Q}^{\mathbf{d}^i}]^l W_{t^*}^\pi = [\mathbf{Q}^{\mathbf{d}^j}]^l W_{t^*}^\pi$ for all $l \geq 0$. Pick $\delta > 0$ sufficiently small such that $\{\mathbf{d}^1, \mathbf{d}^2, \ldots, \mathbf{d}^p\} \subseteq \mathcal{F}_{n+2}(W_t^\pi)$ for all $t \in [t^*, t^* + \delta)$. If the policy π selects $\mathbf{d}^i$ for all $t \in [t^*, t^* + \delta)$, we can write

$$W_t^\pi = e^{[\mathbf{Q}^{\mathbf{d}^i}](t-t^*)} W_{t^*}^\pi \text{ for } t \in [t^*, t^* + \delta),$$

where $e^\Gamma := \sum_{k=0}^\infty \frac{1}{k!}\Gamma^k$ denotes the exponential of a matrix Γ. Therefore, using the fact that $[\mathbf{Q}^{\mathbf{d}^i}]^l W_{t^*}^\pi = [\mathbf{Q}^{\mathbf{d}^j}]^l W_{t^*}^\pi$ for all $l \geq 0$ we have

$$e^{[\mathbf{Q}^{\mathbf{d}^i}](t-t^*)} W_{t^*}^\pi = e^{[\mathbf{Q}^{\mathbf{d}^j}](t-t^*)} W_{t^*}^\pi, \quad \forall\, t \geq t^*. \tag{5}$$

Similarly, we have

$$[\mathbf{Q}^{\mathbf{d}^i}]^l e^{[\mathbf{Q}^{\mathbf{d}^i}]\Delta} W_{t^*}^\pi = [\mathbf{Q}^{\mathbf{d}^j}]^l e^{[\mathbf{Q}^{\mathbf{d}^j}]\Delta} W_{t^*}^\pi, \quad \forall\, l \geq 0 \text{ and } \Delta \geq 0. \tag{6}$$

Now take any $i = \arg\min_j \Delta_j$, thus $\Delta_i \leq \Delta_j$ for all j. Also take $\mathbf{d}' \in \mathcal{F}_{n+2}(W_{t^*+\Delta_i}^\pi)$ and $\mathbf{d}' \neq \mathbf{d}^i$ (this is possible due to the definition of Δ_i). Denote by h the smallest integer for which $1 \leq h \leq n+2$ and

$$[\mathbf{Q}^{\mathbf{d}'}]^h W_{t^*+\Delta_i}^\pi > [\mathbf{Q}^{\mathbf{d}^i}]^h W_{t^*+\Delta_i}^\pi \Rightarrow [\mathbf{Q}^{\mathbf{d}'}]^h e^{[\mathbf{Q}^{\mathbf{d}^i}]\Delta_i} W_{t^*}^\pi > [\mathbf{Q}^{\mathbf{d}^i}]^h e^{[\mathbf{Q}^{\mathbf{d}^i}]\Delta_i} W_{t^*}^\pi.$$

Combining the above expression with Equation (6), we get

$$[\mathbf{Q}^{\mathbf{d}'}]^h e^{[\mathbf{Q}^{\mathbf{d}^i}]\Delta_i} W_{t^*}^\pi > [\mathbf{Q}^{\mathbf{d}^j}]^h e^{[\mathbf{Q}^{\mathbf{d}^j}]\Delta_i} W_{t^*}^\pi \;\Rightarrow\; [\mathbf{Q}^{\mathbf{d}'}]^h W_{t^*+\Delta_i}^\pi > [\mathbf{Q}^{\mathbf{d}^j}]^h W_{t^*+\Delta_i}^\pi,$$

which implies that $\Delta_j \leq \Delta_i$ for any j. The particular selection of i results in $\Delta_j = \Delta_i$ for all i, j. The second part of the proposition is obtained by setting $\Delta = (\delta_2 - \delta_1)$ in Equation (6) and using the definition of the exponential of a matrix. ◀

The above proposition highlights the fact that whenever $\mathcal{F}_{n+2}(W_t^\pi)$ contains more than one decision vector over a time interval, one can construct infinitely many optimal policies by arbitrarily switching between such decision vectors. In the rest of this paper, we restrict our attention to optimal policies that take only mandatory switches: the optimal policy will take an element of $\mathcal{F}_{n+2}(W_t^\pi)$ as long as possible. This does not influence Problems 1 and 2.

The major challenge in the computation of the optimal policy, thus answering the reachability problem, is the computation of the largest time $t \in [0, B)$ such that $\mathcal{F}_{n+2}(W_t^\pi) \neq \mathcal{F}_{n+2}(W_{t^-}^\pi)$, where $W_{t^-}^\pi$ denotes the value of $W_{t-\delta}^\pi$ with δ converging to zero from the right. Suppose a decision vector $\mathbf{d}_0 \in \mathcal{F}_{n+2}(W_0^\pi)$ is selected. The optimal policy will change at the following time point:

$$t'' := \sup \{t \,|\, \mathbf{d}_0 \in \mathcal{F}_{n+2}(W_{t'}^\pi) \text{ for all } t' \in [0, t)\}.$$

3 Conditional Decidability of Problems 1 and 2

3.1 Schanuel's Conjecture and its Implications

Our decidability results will assume Schanuel's Conjecture for the complex numbers, a unifying conjecture in transcendental number theory (see, e.g., [14]). Recall that a *transcendence basis* of a field extension L/K is a subset $S \subseteq L$ such that S is algebraically independent over K and L is algebraic over $K(S)$. The *transcendence degree* of L/K is the (unique) cardinality of some basis.

▶ **Conjecture 7** (**Schanuel's Conjecture (SC)**)**.** *Let $a_1, \ldots, a_n$ be complex numbers that are linearly independent over rational numbers $\mathbb{Q}$. Then the field $\mathbb{Q}(a_1, \ldots, a_n, e^{a_1}, \ldots, e^{a_n})$ has transcendence degree at least n over $\mathbb{Q}$.*

An important consequence of Schanuel's conjecture is that the theory of reals $(\mathbb{R}, 0, 1, +, \cdot, \leq)$ remains decidable when extended with the exponential and trigonometric functions over bounded domains.[1]

▶ **Theorem 8** (Macintyre and Wilkie (see [16, 17]))**.** *Assume **SC**. For any $n \in \mathbb{N}$, the theory $\mathbb{R}_{\mathsf{MW}} := (\mathbb{R}, \exp \restriction [0, n], \sin \restriction [0, n], \cos \restriction [0, n])$ is decidable.*

Our main result will show that Problems 1 and 2 can be decided based on Theorem 8. In fact, Problem 2 can be decided directly from Schanuel's conjecture and recent results on exponential polynomials [11].

▶ **Theorem 9** (Main Result)**.** *Assume **SC**. Then Problems 1 and 2 are decidable.*

In contrast, solving the time-bounded reachability problem for *stationary* policies is decidable unconditionally. This is because fixing a stationary policy reduces the time-bounded reachability problem to one on CTMCs, and one can use the decision procedure from [4].

---

[1] We note that while the result is claimed in several papers [16, 17], a complete proof of this result has never been published. Thus, it would be nice to have a "direct" proof of our main theorem (Theorem 1) starting with Schanuel's conjecture. We do not know such a proof.

3.2 Non-tangential Zeros

Recall that the solution to a first-order linear ODE of dimension n:

$$\frac{d}{dt}X_t = AX_t, \quad z_t = CX_t$$

with real matrices A and C and real initial condition $X_0 \in \mathbb{R}^n$, can be written as $z_t = Ce^{At}X_0$ where e^Γ denotes the exponential of a square matrix Γ, and defined as the infinite sum $e^\Gamma := \sum_{k=0}^{\infty} \frac{1}{k!}\Gamma^k$ that is guaranteed to converge for any matrix Γ. The function can be expressed as an exponential polynomial $z_t = \sum_{j=1}^{k} P_t(j)e^{\lambda_j t}$, where $\lambda_1, \ldots, \lambda_k$ are the distinct (real or complex) eigenvalues of A. Each function $P_t(j)$ is a polynomial function of t possibly with complex coefficients and has a degree one less than the multiplicity of the eigenvalue λ_j. Since the eigenvalues come in conjugate pairs, we can write the real-valued function z as

$$z_t = \sum_{j=1}^{k} e^{a_j t} \sum_{l=0}^{m_j-1} c_{j,l} t^l \cos(b_j t + \varphi_{j,l}), \tag{7}$$

where the eigenvalues are $a_j \pm \mathbf{i}b_j$ with multiplicity m_j. If A, X_0, and C are over the rational numbers, then a_j, b_j, $c_{j,l}$ are real algebraic and $\varphi_{j,l}$ is such that $e^{\mathbf{i}\varphi_{j,l}}$ is algebraic for all j and l. We can symbolically compute derivatives of z which also become functions with a similar closed-form as in (7).

▶ **Definition 10.** *The function z_t has a* zero *at $t = t^*$ if $z_{t^*} = 0$. The zero is said to be* non-tangential *if there is an $\varepsilon > 0$ such that $z_{t_1}z_{t_2} < 0$ for all $t_1 \in (t^* - \varepsilon, t^*)$ and all $t_2 \in (t^*, t^* + \varepsilon)$. The zero is called* tangential *if there is an $\varepsilon > 0$ such that $z_{t_1}z_{t_2} > 0$ for all $t_1 \in (t^* - \varepsilon, t^*)$ and all $t_2 \in (t^*, t^* + \varepsilon)$.*

Note that there are functions with zeros that are neither tangential nor non-tangential. Consider the function $z_t = t\sin\left(\frac{1}{t}\right)$ for $t \neq 0$ and $z_0 = 0$. The function does not satisfy the conditions of being tangential or non-tangential. For any $\varepsilon > 0$, there are $t_1 \in (-\varepsilon, 0)$ and $t_2 \in (0, \varepsilon)$, such that $z_{t_1}z_{t_2} = t_1t_2 \sin\left(\frac{1}{t_1}\right)\sin\left(\frac{1}{t_2}\right)$ is positive. There are also t_1 and t_2 in the respective intervals that make $z_{t_1}z_{t_2}$ negative. In this paper, we only work with functions of the form (7) that are analytic thus infinitely differentiable. Therefore, the first non-zero derivative of z_t at t^* will decide if t^* is tangential or not.

▶ **Proposition 11.** *For any function z_t of the form (7) such that $z_{t^*} = 0$ and $z \not\equiv 0$, there is a k_0 such that $\frac{d^k}{dt^k}z_t\big|_{t=t^*} = 0$ for all $k < k_0$ and $\frac{d^{k_0}}{dt^{k_0}}z_t\big|_{t=t^*} \neq 0$. Moreover, t^* is tangential if k_0 is an even number and is non-tangential if k_0 is an odd number.*

Proof. The proof is based on the Taylor series of z_t at $t = t^*$. Take k_0 the order of the first non-zero derivative of z_t at $t = t^*$. This k_0 always exists since otherwise $z \equiv 0$. The Taylor series of z_t will be

$$z_t = \sum_{k=k_0}^{\infty} \frac{(t-t^*)^k}{k!}\frac{d^k}{dt^k}z_t\big|_{t=t^*} = (t-t^*)^{k_0}\frac{d^{k_0}}{dt^{k_0}}z_t\big|_{t=t^*}\sum_{k=0}^{\infty}\alpha_k(t-t^*)^k, \tag{8}$$

for some $\{\alpha_0, \alpha_1, \ldots\}$ with $\alpha_0 = \frac{1}{k_0!}$. Define the function g by $g_t := \frac{z_t}{(t-t^*)^{k_0}}$ for $t \neq t^*$ and $g_{t^*} := \frac{1}{k_0!}\frac{d^{k_0}}{dt^{k_0}}z_t\big|_{t=t^*}$. Using (8), we get that g is continuous at t^* with $g_{t^*} \neq 0$. Therefore, there is an interval $(t^* - \varepsilon, t^* + \varepsilon)$ over which the function has the same sign as g_{t^*}. For all $t_1 \in (t^* - \varepsilon, t^*)$ and $t_2 \in (t^*, t^* + \varepsilon)$

$$
\begin{aligned}
&g_{t_1} g_{t^*} > 0 \Rightarrow \frac{z_{t_1}}{(t_1 - t^*)^{k_0}} g_{t^*} > 0 \Rightarrow (-1)^{k_0} z_{t_1} g_{t^*} > 0 \\
&g_{t_2} g_{t^*} > 0 \Rightarrow \frac{z_{t_2}}{(t_2 - t^*)^{k_0}} g_{t^*} > 0 \Rightarrow z_{t_2} g_{t^*} > 0 \\
&\Rightarrow (-1)^{k_0} z_{t_1} g_{t^*} z_{t_2} g_{t^*} > 0 \Rightarrow (-1)^{k_0} z_{t_1} z_{t_2} > 0.
\end{aligned}
$$

This means $z_{t_1} z_{t_2} > 0$ for even k_0 and t^* becomes tangential, and $z_{t_1} z_{t_2} < 0$ for odd k_0 and t^* becomes non-tangential. ◀

For any function $z_t = Ce^{At} X_0$, the predicate $\mathsf{NonTangentialZero}(z, l, u)$ stating the existence of a non-tangential zero in an interval (l, u) is expressible in $\mathbb{R}_{\mathsf{MW}}$:

$$\exists t^* . \, l < t^* < u \wedge z_{t^*} = 0 \wedge [\exists \varepsilon > 0 \, . \, \forall t_1 \in (t^* - \varepsilon, 0), t_2 \in (0, t^* + \varepsilon) \, . \, z_{t_1} z_{t_2} < 0]$$

3.3 Switch Points are Non-Tangential Zeroes

Given a CTMDP $\mathcal{M}$ and a piecewise constant optimal policy $\pi : [0, B] \to \mathcal{D}$ for the time-bounded reachability problem, a *switch point* t^* is a point of discontinuity of π. Consider a switch point t^* such that the optimal policy takes the decision vector $\mathbf{d}$ in the time interval $(t^* - \varepsilon)$ and then switches to another decision vector $\mathbf{d}'$ at time t^* for some $\varepsilon > 0$:

$$
\begin{aligned}
&\mathbf{d} \in \mathcal{F}_{n+2}(W_t^\pi) \text{ and } \mathbf{d}' \notin \mathcal{F}_{n+2}(W_t^\pi) \quad \forall t \in (t^* - \varepsilon, t^*), \\
&\mathbf{d} \notin \mathcal{F}_{n+2}(W_t^\pi) \text{ and } \mathbf{d}' \in \mathcal{F}_{n+2}(W_t^\pi) \quad \forall t \in (t^*, t^* + \varepsilon).
\end{aligned}
$$

Consider a (not necessarily unique) state $s \in S$ with actions $a, b \in \mathcal{D}_s$ such that $a \neq b$ and $\mathbf{d}(s) = a$, $\mathbf{d}'(s) = b$. Define the following set of first-order ODEs

$$\Sigma : \begin{cases} \frac{d}{dt} W_t^\pi = \mathbf{Q}^{\mathbf{d}} W_t^\pi \\ z_t = (q^a - q^b) W_t^\pi \end{cases} \tag{9}$$

for $t \in (t^* - \varepsilon, t^* + \varepsilon)$, where q^a and q^b denote the s^{th} row of the matrices $\mathbf{Q}^{\mathbf{d}}$ and $\mathbf{Q}^{\mathbf{d}'}$, respectively. The optimal decision vector on an interval before t^* is $\mathbf{d}$, thus for all $t \in (t^* - \varepsilon, t^*)$,

$$\mathbf{d} \in \mathcal{F}_1(W_t^\pi) \Rightarrow \mathbf{Q}^{\mathbf{d}} W_t^\pi \geq \mathbf{Q}^{\mathbf{d}'} W_t^\pi \Rightarrow (\mathbf{Q}^{\mathbf{d}} - \mathbf{Q}^{\mathbf{d}'}) W_t^\pi \geq 0 \Rightarrow (q^a - q^b) W_t^\pi \geq 0 \Rightarrow z_t \geq 0.$$

The next lemma states that the switch point t^* corresponds to a non-tangential zero for z_t.

▶ **Lemma 12.** *Let π be an optimal piecewise constant policy for the time-bounded reachability problem with bound B. Suppose $\pi(B - t) = \mathbf{d}_t$ for all $t \in [0, B]$. Suppose that for a time point t^*, $\mathbf{d} \in \mathcal{D}$ is an optimal decision before t^* and $\mathbf{d}' \neq \mathbf{d}$ is optimal right after t^*. There is an ε such that for any $s \in S$ with $\mathbf{d}(s) \neq \mathbf{d}'(s)$, $z_t < 0$ for all $t \in (t^*, t^* + \varepsilon)$ with z_t defined in* (9)*.*

Proof. Take k_0 to be the smallest index $k \leq n$ with $\mathbf{d} \notin \mathcal{F}_{k+1}(W_{t^*}^\pi)$ and $\mathbf{d}' \in \mathcal{F}_{k+1}(W_{t^*}^\pi)$. Since $\mathbf{d}'$ is optimal at t^*, we have $\mathbf{d}, \mathbf{d}' \in \mathcal{F}_{k+1}(W_{t^*}^\pi)$ for all $k < k_0$. We show inductively that

$$[\mathbf{Q}^{\mathbf{d}}]^{k+1} W_{t^*}^\pi = [\mathbf{Q}^{\mathbf{d}'}]^{k+1} W_{t^*}^\pi \text{ and } \frac{d^k}{dt^k} z_{t^*} = 0 \text{ for all } 0 \leq k < k_0. \tag{10}$$

The claim is true for $k = 0$:

$$\mathbf{d}, \mathbf{d}' \in \mathcal{F}_1(W^\pi_{t^*}) \Rightarrow \mathbf{Q}^{\mathbf{d}} W^\pi_{t^*} = \mathbf{Q}^{\mathbf{d}'} W^\pi_{t^*}$$
$$\Rightarrow (\mathbf{Q}^{\mathbf{d}} - \mathbf{Q}^{\mathbf{d}'}) W^\pi_{t^*} = \begin{bmatrix} \dots \\ q^a - q^b \\ \dots \end{bmatrix} W^\pi_{t^*} = 0 \Rightarrow (q^a - q^b) W^\pi_{t^*} = 0 \Rightarrow z_{t^*} = 0.$$

Now suppose (10) holds for $(k-1)$ with $k < k_0$. Then

$$\mathbf{d}, \mathbf{d}' \in \mathcal{F}_{k+1}(W^\pi_{t^*}) \Rightarrow [\mathbf{Q}^{\mathbf{d}}]^{k+1} W^\pi_{t^*} = [\mathbf{Q}^{\mathbf{d}'}]^{k+1} W^\pi_{t^*}$$
$$\Rightarrow \mathbf{Q}^{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^{k} W^\pi_{t^*} = \mathbf{Q}^{\mathbf{d}'}[\mathbf{Q}^{\mathbf{d}'}]^{k} W^\pi_{t^*} \Rightarrow^{(*)} \mathbf{Q}^{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^{k} W^\pi_{t^*} = \mathbf{Q}^{\mathbf{d}'}[\mathbf{Q}^{\mathbf{d}}]^{k} W^\pi_{t^*}$$
$$\Rightarrow [\mathbf{Q}^{\mathbf{d}} - \mathbf{Q}^{\mathbf{d}'}][\mathbf{Q}^{\mathbf{d}}]^{k} W^\pi_{t^*} = 0 \Rightarrow^{(**)} [\mathbf{Q}^{\mathbf{d}} - \mathbf{Q}^{\mathbf{d}'}] \frac{d^k}{dt^k} X_{t^*} = 0$$
$$\Rightarrow (q^a - q^b) \frac{d^k}{dt^k} X_{t^*} = 0 \Rightarrow \frac{d^k}{dt^k} z_{t^*} = 0,$$

where $(*)$ holds due to the induction assumption and $(**)$ is true due to the differential equation (9). Finally, we show that $\frac{d^{k_0}}{dt^{k_0}} z_{t^*} < 0$.

$$\mathbf{d} \notin \mathcal{F}_{k_0+1}(W^\pi_{t^*}) \text{ and } \mathbf{d}' \in \mathcal{F}_{k_0+1}(W^\pi_{t^*}) \Rightarrow [\mathbf{Q}^{\mathbf{d}}]^{k_0+1} W^\pi_{t^*} < [\mathbf{Q}^{\mathbf{d}'}]^{k_0+1} W^\pi_{t^*}$$
$$\Rightarrow \mathbf{Q}^{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^{k_0} W^\pi_{t^*} < \mathbf{Q}^{\mathbf{d}'}[\mathbf{Q}^{\mathbf{d}'}]^{k_0} W^\pi_{t^*} \Rightarrow^{(\imath)} \mathbf{Q}^{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^{k_0} W^\pi_{t^*} < \mathbf{Q}^{\mathbf{d}'}[\mathbf{Q}^{\mathbf{d}}]^{k_0} W^\pi_{t^*}$$
$$\Rightarrow [\mathbf{Q}^{\mathbf{d}} - \mathbf{Q}^{\mathbf{d}'}] \frac{d^{k_0}}{dt^{k_0}} W^\pi_{t^*} < 0 \Rightarrow (q^a - q^b) \frac{d^{k_0}}{dt^{k_0}} W^\pi_{t^*} < 0 \Rightarrow \frac{d^{k_0}}{dt^{k_0}} z_{t^*} < 0,$$

where $(\imath)$ holds due to (10) for $k_0 - 1$.

Since $z_{t^*} = 0$, we can select ε such that $z_t > 0$ for all $t \in (t^* - \varepsilon, t^*)$. Using Taylor expansion (8) and the facts that $\frac{d^{k_0}}{dt^{k_0}} z_{t^*} < 0$ and $z_t > 0$ for $t \in (t^* - \varepsilon, t^*)$, we have that k_0 must be an odd number, which means t^* is non-tangential by Prop. 11. The function z_t changes sign from positive to negative at t^*. ◀

3.4 Conditional Decidability

The decision procedure for Problem 1 is as follows. Fix a CTMDP $\mathcal{M} = (\{1, \dots, n\} \uplus \{\mathbf{good}\}, \mathcal{D}, \mathbf{Q})$ and a bound B. We inductively construct a piecewise constant optimal policy, going forward in time. To begin, we set the initial decision vector to $\mathbf{d}^1$, where $\mathbf{d}^1$ is selected such that $\mathbf{d}^1 \in \mathcal{F}_{n+2}(W^\pi_0)$ (Equation (4)) with W^π_0 set to the indicator vector that is 1 at the **good** state and 0 in other states.

Note that in general $\mathcal{F}_{n+2}(W^\pi_t)$ in (4) may have finitely many elements and the choice of optimal decision at time t, $\mathbf{d}_t \in \mathcal{F}_{n+2}(W^\pi_t)$ is not unique. Based on results of Proposition 6, any arbitrary element of $\mathcal{F}_{n+2}(W^\pi_t)$ can be chosen; but, we do not alter this choice until the picked decision vector does not belong to $\mathcal{F}_{n+2}(W^\pi_t)$ anymore. We know that there is a piecewise constant optimal policy π with finitely many switches obtained from the charactrization in Theorem 5. Denote the (unknown) number of switches by $k \in \mathbb{N}$.

We find k as follows. We inductively check the existence of a sequence of decision vectors $\mathbf{d}^1, \dots, \mathbf{d}^k$ and time points $t_1, \dots, t_{k-1}$ such that the optimal policy (given a lexicographical order on $\mathcal{D}$) switches from $\mathbf{d}^i$ to $\mathbf{d}^{i+1}$ at time t_i but does not have any switch between the time points. Then, we check if the optimal policy makes at least one additional switch point in the interval (t_k, B). The check reduces the question to a number of satisfiability questions in $\mathbb{R}_{\mathsf{MW}}$. If we find an additional switch, we know that the optimal strategy has at least $k+1$ switches and continue to check if there are further switch points. If not, we know that the optimal policy has k switch points.

We need some notation. A *prefix* $\sigma_k = (\mathbf{d}^1, t_1, \mathbf{d}^2, t_2, \ldots, t_{k-1}, \mathbf{d}^k) \in (\mathcal{D} \times (0, B))^* \times \mathcal{D}$ is a finite sequence of decision vectors from $\mathcal{D}$ and strictly increasing time points $0 < t_1 < t_2 < \ldots < t_{k-1} < B$ such that $\mathbf{d}^i \neq \mathbf{d}^{i+1}$ for $i \in \{1, \ldots, k-1\}$. Intuitively, it represents the prefix of a piecewise constant policy with the first $k-1$ switches. For two decision vectors $\mathbf{d}, \mathbf{d}'$, let $\Delta(\mathbf{d}, \mathbf{d}') := \{s \mid \mathbf{d}(s) \neq \mathbf{d}'(s)\}$ be the states at which the actions suggested by the decision vectors differ. For a decision vector $\mathbf{d}$, let $\mathbf{d}[s \mapsto b]$ denote the decision vector that maps state s to action b but agrees with $\mathbf{d}$ otherwise.

For a prefix $\sigma_k = (\mathbf{d}^1, t_1, \mathbf{d}^2, t_2, \ldots, t_{k-1}, \mathbf{d}^k)$, a state $s \in S$, and an action $b \in \mathcal{D}_s$, define

$$y_t^{s,b}(\sigma_k) = \mathbf{u}^T(s)([\mathbf{Q}^{\mathbf{d}^k}] - [\mathbf{Q}^{\mathbf{d}^k[s \mapsto b]}])e^{[\mathbf{Q}^{\mathbf{d}^k}](t-t_{k-1})}e^{[\mathbf{Q}^{\mathbf{d}^{k-1}}](t_{k-1}-t_{k-2})} \cdots e^{[\mathbf{Q}^{\mathbf{d}^1}]t_1}\mathbf{u}(\mathbf{good}),$$

where $\mathbf{u}(s)$ is a vector of dimension $n+1$ that assigns one to s and zero to every other entry. Observe that $y_t^{s,b}(\sigma_k)$ is a solution of a set of linear ODEs similar to z_t in Equation (9):

$$\begin{cases} \frac{d}{dt}W_t = [\mathbf{Q}^{\mathbf{d}^k}]W_t \\ y_t^{s,b}(\sigma_k) = \mathbf{u}^T(s)([\mathbf{Q}^{\mathbf{d}^k}] - [\mathbf{Q}^{\mathbf{d}^k[s \mapsto b]}])W_t, \end{cases} \tag{11}$$

with the condition $W_{t_{k-1}} = e^{[\mathbf{Q}^{\mathbf{d}^{k-1}}](t_{k-1}-t_{k-2})} \cdots e^{[\mathbf{Q}^{\mathbf{d}^1}]t_1}\mathbf{u}(\mathbf{good})$.

We shall use (variants of) the predicate $\mathsf{NonTangentialZero}(y^{\cdot,\cdot}, t_1, t_2)$, but write the predicates informally for readability. We need two additional predicates $\mathsf{Switch}(\sigma_k, t^*, \mathbf{d}')$ and $\mathsf{NoSwitch}(\sigma_{k+1})$. The predicate Switch states that, given a prefix σ_k, the first switch from $\mathbf{d}^k$ to a new decision vector $\mathbf{d}'$ occurs at time point $t^* > t_{k-1}$. This new switch requires three conditions. First, there is a simultaneous non-tangential zero at t^* for all dynamical systems of the form (11) associated with $y_t^{s,\mathbf{d}'(s)}(\sigma_k)$, $s \in \Delta(\mathbf{d}^k, \mathbf{d}')$. Second, t^* is the first time after t_{k-1} that any of the dynamical systems have a non-tangential zero. Finally, none of the states in $S \setminus \Delta(\mathbf{d}^k, \mathbf{d}')$ whose action remains the same before and after the switch, have a non-tangential zero in $(t_{k-1}, t^*]$ (up to and including t^*):

$$\begin{aligned} &\mathsf{Switch}(\underbrace{(\mathbf{d}^1, t_1, \ldots, t_{k-1}, \mathbf{d}^k)}_{\sigma_k}, t^*, \mathbf{d}') \equiv \\ &\quad 0 < t_1 < \ldots < t_{k-1} < B \wedge (B > t^* > t_{k-1}) \wedge (\Delta(\mathbf{d}^k, \mathbf{d}') \neq \emptyset) \wedge \\ &\quad \bigwedge_{s \in \Delta(\mathbf{d}^k, \mathbf{d}')} \begin{pmatrix} \text{“}y_t^{s,\mathbf{d}'(s)}(\sigma_k) \text{ has a non-tangential zero at } t^*\text{”} \wedge \\ \text{“}y_t^{s,\mathbf{d}'(s)}(\sigma_k) \text{ has no non-tangential zero in } (t_{k-1}, t^*)\text{”} \end{pmatrix} \wedge \\ &\quad \bigwedge_{s \in S \setminus \Delta(\mathbf{d}^k, \mathbf{d}')} \text{“}y_t^{s,\mathbf{d}'(s)}(\sigma_k) \text{ has no non-tangential zero in } (t_{k-1}, t^*]\text{”} \end{aligned}$$

The predicate $\mathsf{NoSwitch}(\sigma_{k+1})$ states that, given a prefix σ_{k+1}, the last decision vector $\mathbf{d}^{k+1}$ of (σ_{k+1}) stays optimal and does not switch to another decision vector within the interval (t_k, B). This is equivalent to stating that none of the dynamical systems of the from (11) associated with $y_t^{s,b}(\sigma_{k+1})$ for $s \in S, b \in \mathcal{D}_s \setminus \mathbf{d}^{k+1}(s)$ has a non-tangential zero in (t_k, B):

$$\mathsf{NoSwitch}(\sigma_{k+1}) \equiv \bigwedge_{s, b \neq \mathbf{d}^{k+1}(s)} \text{“}y_t^{s,b}(\sigma_{k+1}) \text{ has no non-tangential zero in } (t_k, B)\text{”}$$

We can now check if the optimal strategy has exactly k switches. The first part of the predicate written below sets up a proper σ and the last conjunct states that there is no further switch after the last one.

$$\exists t_1, \ldots, t_k.(0 < t_1 < t_2 \ldots < t_k < B) \wedge \bigwedge_{i=1}^{k} \mathsf{Switch}(\underbrace{\mathbf{d}^1, t_1, \ldots, \mathbf{d}^i, t_i, \mathbf{d}^{i+1}}_{\sigma_{i+1}}) \wedge \mathsf{NoSwitch}(\sigma_{k+1}).$$

We can enumerate these formulas with increasing k over all choices of decision vectors and stop when the above formula is valid. At this point, we know that there is a piecewise constant optimal policy with k switches, which plays the decision vectors $\mathbf{d}^1, \ldots, \mathbf{d}^k$. We can make one more query to check if the probability of reaching **good** when playing this strategy is at least a given rational vector $r \in [0,1]^n$:

$$\exists t_1, \ldots, t_k.(0 < t_1 < \ldots, t_k < B) \wedge \bigwedge_{i=1}^{k} \mathsf{Switch}(\mathbf{d}^1, t_1, \ldots, \mathbf{d}^i, t_i, \mathbf{d}^{i+1}) \wedge \mathsf{NoSwitch}(\sigma_{k+1})$$
$$\wedge \bigwedge_{s=1}^{n} \mathbf{u}^T(s) e^{[\mathbf{Q}^{\mathbf{d}^{k+1}}](B-t_k)} e^{[\mathbf{Q}^{\mathbf{d}^k}](t_k - t_{k-1})} \ldots e^{[\mathbf{Q}^{\mathbf{d}^1}]t_1} \mathbf{u}(\mathbf{good}) > r(s) \tag{12}$$

This completes the proof of conditional decidability of Problem 1.

Conditional Decidability for Problem 2. A stationary policy $\mathbf{d}$ is not optimal if there is a switch point. Using the Switch predicate and conditional decidability of $\mathbb{R}_{\mathsf{MW}}$, this shows conditional decidability of Problem 2.

In fact, to check the presence of a single non-tangential zero, one can avoid Theorem 8 and get a direct construction based on Schanuel's conjecture. This construction is similar to [11] and is provided in Section 5. Unfortunately, when there are multiple switch points, we have to existentially quantify over previous switch points. Thus, the techniques of [11] cannot be straightforwardly extended to find a direct conditional decision procedure for Problem 1.

We do not know if there is a numerical procedure that only uses an oracle for non-tangential zeros. The problem is that, while numerical techniques can be used to bound each non-tangential zero with rational intervals with arbitrary precision as well as compute the reachability probability to arbitrary precision, we do not know how to numerically detect whether the reachability probability in (12) is exactly equal to a given r. By the Lindemann-Weierstrass Theorem [15], we already know that for CTMDPs with stationary optimal strategies, the value of reachability probability for any rational time bound $B > 0$ is transcendental and hence $\sup_{\pi \in \Pi_B} \mathbf{P}_s^{\pi}(\mathbf{reach}) \neq r(s)$ for all $s \in S$. However, we cannot prove that the reachability probability remains irrational in the general case.

4 Lower Bound: Continuous Skolem Problem

▶ **Problem 3** (Bounded Continuous-Time Skolem Problem)**.** *Given a linear ordinary differential equation (ODE)*

$$\frac{d^n}{dt^n} z_t + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} z_t + \cdots + a_1 \frac{d}{dt} z_t + a_0 z_t = 0 \tag{13}$$

with rational initial conditions $z_0, \frac{dz_t}{dt}|_{t=0}, \ldots, \frac{d^{n-1}z_t}{dt^{n-1}}|_{t=0} \in \mathbb{Q}$ *and rational coefficients* $a_{n-1}, a_{n-2}, \ldots, a_0 \in \mathbb{Q}$ *and a time bound* $B \in \mathbb{Q}$*, the* bounded continuous Skolem problem *asks whether there exists* $0 < t^* < B$ *such that it is a non-tangential zero for* z_t*. Further, we can assume w.l.o.g. that* $z_0 = 0$ *in the initial condition.*[2]

[2] The assumption is w.l.o.g. because given a linear ODE whose solution is z_t, one can construct another linear ODE whose solution is $y_t = t z_t$. Clearly, $y_0 = 0$ and there is a non-tangential zero of z in $(0, B)$ iff there is a non-tangential zero of y in $(0, B)$.

We note that our definition is slightly different from the usual definition of the problem, e.g., in [7, 11], which simply asks for any zero (i.e., $z_{t^*} = 0$), not necessarily a non-tangential one. Our version of the bounded continuous Skolem problem is also decidable assuming **SC** [11]. However, there is no unconditional decidability result known for this problem, even though we only look for a non-tangential zero.

We can encode any given linear ODE of order n in the form of (13) into a set of n first-order linear ODE on $X : [0, B] \to \mathbb{R}^n$ with

$$\begin{cases} \frac{d}{dt} X_t = AX_t, \quad X_0 = \left[z_0, \left.\frac{dz_t}{dt}\right|_{t=0}, \dots, \left.\frac{d^{n-1}z_t}{dt^{n-1}}\right|_{t=0} \right]^T \\ z_t = CX_t, \end{cases} \tag{14}$$

with the state matrix A and output matrix C are

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{15}$$

Using the representation (14), the solution of the linear ODE (13) can be written as $z_t = Ce^{At}X_0$. Therefore, the bounded continuous-time Skolem problem can be restated as whether the expression $Ce^{At}X_0$ has a non-tangential zero in the interval $(0, B)$.

We now reduce the bounded continuous-time Skolem problem to Problem 2. Given an instance (14)-(15) of the Skolem problem of dimension n, we shall construct a CTMDP over states $\{1, \dots, 2n\} \cup \{\mathbf{good}, \mathbf{bad}\}$ and bound B, and just two decision vectors $\mathbf{d}^a$ and $\mathbf{d}^b$ that only differ in the available actions (a or b) at state 1. Our reduction will ensure that the answer of the Skolem problem has a non-tangential zero iff there is a switch in the optimal policy in the time-bounded reachability problem for bound B, and thus, iff stationary policies are not optimal.

▶ **Theorem 13.** *For every instance of the bounded continuous-time Skolem problem with dynamics $\frac{d}{dt}X_t = AX_t$, $z_t = CX_t$, initial condition X_0, and time bound B, there is a CTMDP $\mathcal{M}$ such that the dynamical system has a non-tangential zero in $(0, B)$ iff the optimal strategy of the CTMDP in the time-bounded reachability problem is not stationary.*

We sketch the main ideas of the proof here. Consider the linear differential equation described by the state space representation in (14) with the initial condition X_0 that has its first element equal to zero $X_0(1) = 0$. Given the time bound $B > 0$, to solve the bounded continuous Skolem problem, we are looking for the existence of a time $0 < t^* < B$ such that $z_{t^*} = 0$ is non-tangential. Equivalently, we want to find a non-tangential zero for the function $Ce^{At}X_0$, where $C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$.

There are three obstacles to go from (14) to generator matrices for a CTMDP. Each generator matrix must have non-diagonal entries that are non-negative. The sum of each row of the matrix must be zero. Moreover, the last state of the CTMDP must be absorbing. None of these properties may hold for a general A. We show a series of transformations that take the matrix A to a matrix P that is sub-stochastic. Then we construct the generator matrices of the CTMDP using P that include the required absorbing state. We denote by $\mathbf{0}_{\mathsf{m}}$ and $\mathbf{1}_{\mathsf{m}}$ as row vectors of dimension m with all elements equal to zero and one, respectively.

▶ **Theorem 14.** *Suppose $A \in \mathbb{Q}^{n\times n}$, $X_0 \in \mathbb{Q}^n$ and $C = [1, \mathbf{0}_{n-1}]$ are given with $X_0(1) = 0$. There are positive constants γ, λ and a* generator *matrix $P \in \mathbb{Q}^{(2n+1)\times(2n+1)}$ such that*

$$Ce^{At}X_0 = \gamma e^{\lambda t}\left[C'e^{Pt}Y_0\right], \quad C' = [1, -1, \mathbf{0}_{2n-1}], \quad Y_0 = [\mathbf{0}_{2n}, 1]^T. \tag{16}$$

▶ Remark 15. The first equality in (16) ensures that nature of zeros of the two functions $Ce^{At}X_0$ and $C'e^{Pt}Y_0$ are the same. If one of them has a non-tangential zero at t^* the other one will also have a non-tangential zero at t^*. To see this, suppose $Ce^{At^*}X_0 = 0$ and $Ce^{At}X_0$ changes sign at t^*. The same things happen to $C'e^{Pt}Y_0$ due to the fact that the two functions are different with only a positive factor of $\gamma e^{\lambda t}$.

Without loss of generality, we assume the element A_{11} is negative. This assumption is needed when constructing the CTMDP in the sequel. If the assumption does not hold, we can always replace A with $A - \lambda_0\mathbb{I}_n$ for a sufficiently large λ_0 and merge λ_0 with λ in (16). Define the map $\phi_1 : \cup_n\mathbb{Q}^{n\times n} \to \cup_n\mathbb{Q}^{2n\times 2n}_{\geq 0}$ such that $\phi_1(A)$ is obtained by replacing each entry A_{ij} with the matrix $\begin{bmatrix}\alpha_{ij} & \beta_{ij}\\ \beta_{ij} & \alpha_{ij}\end{bmatrix}$, where $\alpha_{ij} = max(A_{ij}, 0)$ and $\beta_{ij} = max(-A_{ij}, 0)$. The map ϕ_1 maps any square matrix to another matrix with non-negative entries ([2]). Also define the map $\phi_2 : \cup_n\mathbb{Q}^n \to \cup_n\mathbb{Q}^{2n}$ such that $\phi_2(X)$ replaces each entry $X(i)$ with two entries $[X(i), 0]^T$.

▶ **Proposition 16.** *We have $C''e^{\phi_1(A)t}Y_2 = Ce^{At}X_0$ with $Y_2 := \phi_2(X_0)$ and $C'' := [1, -1, \mathbf{0}_{2n-2}]$.*

Proof. We can show inductively that for any $k \in \{0, 1, 2, \ldots\}$, $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, and $[\beta_1, \beta_2, \ldots, \beta_n] := [\alpha_1, \alpha_2, \ldots, \alpha_n]A^k$, we have

$$[\alpha_1, -\alpha_1, \alpha_2, -\alpha_2, \ldots, \alpha_n, -\alpha_n]\phi_1(A)^k = [\beta_1, -\beta_1, \beta_2, -\beta_2, \ldots, \beta_n, -\beta_n].$$

Substitute $[\alpha_1, \alpha_2, \ldots, \alpha_n]$ by C and $[\beta_1, \beta_2, \ldots, \beta_n] = CA^k$ to get

$$\begin{aligned}
&C''\phi_1(A)^kY_2 = C''\phi_1(A)^k\phi_2(X_0) = [\beta_1, -\beta_1, \beta_2, -\beta_2, \ldots, \beta_n, -\beta_n]\phi_2(X_0)\\
&= [\beta_1, \beta_2, \ldots, \beta_n]X_0 = CA^kX_0\\
&\Rightarrow C''e^{\phi_1(A)t}Y_2 = \sum_{k=0}^{\infty}\frac{t^k}{k!}C''\phi_1(A)^kY_2 = \sum_{k=0}^{\infty}\frac{t^k}{k!}CA^kX_0 = Ce^{At}X_0.
\end{aligned}$$ ◀

Next, we define $\lambda := \max_i \sum_{j=1}^n |A_{ij}| + 1$, $P_2 := \phi_1(A) - \lambda\mathbb{I}_n$, and the vector $\boldsymbol{\beta} \in \mathbb{Q}^{2n}$ with

$$\boldsymbol{\beta}(2i-1) = \boldsymbol{\beta}(2i) = \max(0, -P_2Y_2(2i-1), -P_2Y_2(2i)) \quad 1 \leq i \leq n.$$

Note that the row sum of P_2 is at most -1 and $\boldsymbol{\beta} + P_2Y_2$ is element-wise non-negative with the maximum element

$$\gamma := \max_i P_2Y_2(i) + \boldsymbol{\beta}(i).$$

▶ **Proposition 17.** *The above choices of λ, γ and the matrix*

$$P := \begin{bmatrix} P_2 & \vdots & (P_2Y_2 + \boldsymbol{\beta})/\gamma\\ \cdots & \cdots & \cdots\\ \mathbf{0} & \vdots & 0\end{bmatrix}$$

satisfy (16) in Theorem 14. Moreover, P is row sub-stochastic.

Proof. We can easily show by induction that

$$P^k Y_0 = \begin{bmatrix} P_2^{k-1}(P_2 Y_2 + \boldsymbol{\beta})/\gamma \\ 0 \end{bmatrix}, \quad \forall k \in \{1, 2, \ldots\}.$$

$$C' e^{Pt} Y_0 = \sum_{k=0}^{\infty} \frac{t^k}{k!} C' P^k Y_0 = C' Y_0 + C'' \sum_{k=1}^{\infty} \frac{t^k}{k!} P_2^{k-1}(P_2 Y_2 + \boldsymbol{\beta})/\gamma,$$

where $C'' := [1, -1, \mathbf{0}_{2n-2}]$ is the same vector as C' but the last element is eliminated.

$$C' e^{Pt} Y_0 = C' Y_0 + C'' e^{P_2 t} Y_2/\gamma - C'' Y_2/\gamma + \sum_{k=1}^{\infty} \frac{t^k}{k!} C'' P_2^{k-1} \boldsymbol{\beta}/\gamma.$$

The term $C' Y_0$ is zero by simple multiplication of the two vectors. $C'' Y_2 = C'' \phi_2(X_0) = X_0(1)$, which is also assumed to be zero. Finally, we see by induction that for all $k \in \{1, 2, \ldots\}$, the elements $(2i-1)$ and $2i$ of the matrix $P_2^{k-1}\boldsymbol{\beta}$ are equal due to the particular structure of P_2 and $\boldsymbol{\beta}$. Therefore, the last sum in the above is also zero and we get

$$\begin{aligned} C' e^{Pt} Y_0 &= C'' e^{P_2 t} Y_2/\gamma = C'' e^{\phi_1(A)t - \lambda \mathbb{I} t} \phi_2(X_0)/\gamma \\ &= C'' e^{\phi_1(A)t} \phi_2(X_0) e^{-\lambda t}/\gamma = C e^{At X_0} e^{-\lambda t}/\gamma. \end{aligned}$$

To show that P is a sub-stochastic matrix, we recall that $P_2 Y_2 + \boldsymbol{\beta} \geq 0$ with maximum element γ. Then

$$P_2 \times \mathbf{1}_{2n} + (P_2 Y_2 + \boldsymbol{\beta})/\gamma \leq \phi_1(A)\mathbf{1}_{2n} - \lambda \mathbf{1}_{2n} + \mathbf{1}_{2n} = \phi_1(A)\mathbf{1}_{2n} - \max_i \sum_j |A_{ij}| \leq 0. \quad \blacktriangleleft$$

As the last step, we add an additional row and column to P to make it stochastic:

$$\mathbf{Q}^a := \begin{bmatrix} P_2 & \vdots & \Theta & \vdots & (P_2 Y_2 + \boldsymbol{\beta})/\gamma \\ \cdots & & \cdots & & \cdots \\ \mathbf{0}_{2\times 2n} & \vdots & \mathbf{0}_{2\times 1} & \vdots & \mathbf{0}_{2\times 1} \end{bmatrix}, \bar{C} = \begin{bmatrix} 1 & -1 & \mathbf{0}_{2n} \end{bmatrix}, \bar{Y}_0 = \begin{bmatrix} \mathbf{0}_{2n+1} \\ 1 \end{bmatrix},$$

where Θ has non-negative entries and is such that $\mathbf{Q}^a$ is stochastic (sum of elements of each row is zero). The added row and column correspond to an absorbing state for a CTMDP with no effect on reachability probability: $\bar{C} e^{t\mathbf{Q}^a} \bar{Y}_0 = C' e^{Pt} Y_0$.

Next, we obtain a second generator matrix for the CTMDP. Define $\mathbf{Q}^b := \mathbf{Q}^a + K$ with

$$K := \begin{bmatrix} -r & r & \mathbf{0}_{2n} \\ \mathbf{0}_{(2n+1)\times 1} & \mathbf{0}_{(2n+1)\times 1} & \mathbf{0}_{(2n+1)\times 2n} \end{bmatrix},$$

Note that $\mathbf{Q}^b$ has exactly the same transition rates as in $\mathbf{Q}^a$ except the transition from state 1 to state 2, which is changed by r.

▶ Remark 18. We assumed w.l.o.g. that A_{11} is negative. The construction of $P_2, P, \mathbf{Q}^a$ results in a positive value for $\mathbf{Q}^a_{12}$. Therefore, it is possible to select both negative and positive values for r such that $\mathbf{Q}^b_{12} = \mathbf{Q}^a_{12} + r \geq 0$.

Construction of the CTMDP. The CTMDP $\mathcal{M}$ has $2n+2$ states, corresponding to the rows of $\mathbf{Q}^a$ and $\mathbf{Q}^b$, with the absorbing state $2n+2$ associated with the **good** state and the absorbing state $2n+1$ with reachability probability equal to zero. We shall set the time bound to be B. $\mathcal{D}_s$ the set of actions that can be taken in state $s \in \{2, 3, \ldots, 2n+2\}$

is singleton and $\mathcal{D}_1 = \{a, b\}$. The set of decision vectors has two elements $\mathcal{D} = \{\mathbf{d}^a, \mathbf{d}^b\}$ corresponding to the actions a, b taken at state 1. For simplicity, we denote the generator matrices of these decision vectors by $\mathbf{Q}^a$ and $\mathbf{Q}^b$, respectively. Moreover, the two actions a, b have the same transition rates for jumping from state 1 to other states, except giving different rates r_a, r_b for jumping from 1 to 2 such that $r_b - r_a = r$.

The optimal policy π takes decision vector $\mathbf{d}_t \in \mathcal{D}$ at time $B - t$ such that $\mathbf{d}_t \in \mathcal{F}_{n+2}(W_t^\pi)$ for all $t \in [0, B]$ as defined in (4).

▶ **Proposition 19.** *Let r have the same sign of the first non-zero element of the set $\{\bar{C}\bar{Y}_0, \bar{C}\mathbf{Q}^a\bar{Y}_0, \bar{C}(\mathbf{Q}^a)^2\bar{Y}_0, \ldots\}$ and such that $\mathbf{Q}^a_{12} + r \geq 0$. This particular selection of r results in the optimality of $\mathbf{d}^a$ at $t = 0$.*

Proof. We have $W_0^\pi = \bar{Y}_0$ and $\mathcal{F}_k(W_0^\pi) = \arg\max_{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^k\bar{Y}_0$. Then, we need to compare $[\mathbf{Q}^a]^k\bar{Y}_0$ with $[\mathbf{Q}^b]^k\bar{Y}_0$ for different values of k and see which one gives the first highest value. These two are the same for $k = 1$ and $\mathcal{F}_1(W_0^\pi) = \arg\max_{\mathbf{d}} \mathbf{Q}^{\mathbf{d}}\bar{Y}_0 = \{\mathbf{d}^a, \mathbf{d}^b\}$. Suppose For $k_0 > 1$ is the smallest index such that $\bar{C}[\mathbf{Q}^a]^{k_0}\bar{Y}_0 \neq 0$. It can be shown inductively that $[\mathbf{Q}^b]^k\bar{Y}_0 = [\mathbf{Q}^a]^k\bar{Y}_0$ for all $1 \leq k \leq k_0$:

$$\begin{aligned}
[\mathbf{Q}^b]^k\bar{Y}_0 &= \mathbf{Q}^b[\mathbf{Q}^b]^{k-1}\bar{Y}_0 = (\mathbf{Q}^a + K)[\mathbf{Q}^b]^{k-1}\bar{Y}_0 = (\mathbf{Q}^a + K)[\mathbf{Q}^a]^{k-1}\bar{Y}_0 \\
&= [\mathbf{Q}^a]^k\bar{Y}_0 + K[\mathbf{Q}^a]^{k-1}\bar{Y}_0 = [\mathbf{Q}^a]^k\bar{Y}_0 - r\begin{bmatrix}\bar{C}[\mathbf{Q}^a]^{k-1}\bar{Y}_0 \\ \mathbf{0}_{(2n+1)\times 2n}\end{bmatrix} = [\mathbf{Q}^a]^k\bar{Y}_0.
\end{aligned}$$

This means $\mathcal{F}_k(W_0^\pi) = \arg\max_{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^k\bar{Y}_0 = \{\mathbf{d}^a, \mathbf{d}^b\}$ for all $1 \leq k \leq k_0$. We have for $k = k_0 + 1$

$$[\mathbf{Q}^b]^{k_0+1}\bar{Y}_0 = [\mathbf{Q}^a]^{k_0+1}\bar{Y}_0 - r\begin{bmatrix}\bar{C}[\mathbf{Q}^a]^{k_0}\bar{Y}_0 \\ \mathbf{0}_{(2n+1)\times 2n}.\end{bmatrix}$$

The first element of $[\mathbf{Q}^b]^{k_0+1}\bar{Y}_0$ is strictly less than the first element of $[\mathbf{Q}^a]^{k_0+1}\bar{Y}_0$ since r has the same sign as $\bar{C}[\mathbf{Q}^a]^{k_0}\bar{Y}_0$. Thus $\mathcal{F}_{k_0+1}(W_0^\pi) = \arg\max_{\mathbf{d}}[\mathbf{Q}^{\mathbf{d}}]^{k_0+1}\bar{Y}_0 = \{\mathbf{d}^a\}$. ◀

Note that the Skolem problem is trivial with the solution $z_t = 0$ for all $t \in [0, B]$ if all the elements of the set $\{\bar{C}\bar{Y}_0, \bar{C}\mathbf{Q}^a\bar{Y}_0, \bar{C}(\mathbf{Q}^a)^2\bar{Y}_0, \ldots\}$ are zero.

Prop. 19 guarantees existence of an $\varepsilon \in (0, B)$ such that W_t^π satisfies

$$\frac{d}{dt}W_t^\pi = \mathbf{Q}^a W_t^\pi \quad \forall t \in (0, \varepsilon),$$

with the initial condition $W_0^\pi(2n+2) = 1$ and $W_0^\pi(s) = 0$ for all $s \in \{1, 2, \ldots, 2n+1\}$.

To check if the optimal policy switches to $\mathbf{d}^b$ at some time point, we should check if there is $t^* < B$ such that $\mathbf{d}^b \in \mathcal{F}_{n+2}(W_{t^*}^\pi)$. This is equivalent to having t^* being non-tangential for the maximization in $\mathcal{F}_1(W_t^\pi)$, which means t^* is non-tangential for the equation

$$\mathbf{Q}^a W_t^\pi = \mathbf{Q}^b W_t^\pi \Leftrightarrow K W_t^\pi = 0 \Leftrightarrow \bar{C} W_t^\pi = 0.$$

Summarizing the above derivations, we have the following set of ODEs

$$\frac{d}{dt}W_t^\pi = \mathbf{Q}^a W_t^\pi \quad, W_0^\pi = \bar{Y}_0, \quad z_t = \bar{C} W_t^\pi. \tag{17}$$

The optimal policy for CTMDP $\mathcal{M}$ switches from $\mathbf{d}^a$ to $\mathbf{d}^b$ at some time point t^* if and only if z_t in (17) has a non-tangential zero in $(0, B)$ if and only if the original dynamics $Ce^{At}X_0$ has a non-tangential zero in $(0, B)$. This completes the proof of Theorem 13.

5 A Direct Algorithm for Problem 2

We now show a "direct" method for decidability of Problem 2 based on Schanuel's conjecture but without relying on the decidability of $\mathbb{R}_{\mathsf{MW}}$. As stated before, a switch point in a strategy corresponds to the existence of a non-tangential zero for the functions $y_t^{s,b}(\mathbf{d}^1)$ for $s \in S$ and $b \in \mathcal{D}_s \setminus \mathbf{d}^1(s)$. We know $y_t^{s,b}(\mathbf{d}^1)$ is an exponential polynomial of the form (7). Thus, deciding Problem 2 reduces to checking if an exponential polynomial of the form (7) in one free variable t has a non-tangential zero in a bounded interval. We use the following result from [11].

▶ **Theorem 20** ([11]). *Assume **SC**. It is decidable whether an exponential polynomial of the form* (7) *has a zero in the interval* (t_1, t_2) *with* $t_1, t_2 \in \mathbb{Q}$.

Theorem 20 decides whether a zero, not necessarily a non-tangential one, exists. We shall use the characterization of Proposition 11 to check if a non-tangential zero of $y_t := y_t^{s,b}(\mathbf{d}^1)$ exists in $(0, B)$. Define the functions

$$z_t^k = y_t^2 + \sum_{j=1}^{k}\left(\frac{d^j}{dt^j}y_t\right)^2, \quad k \in \{0,1,2,\ldots\}. \tag{18}$$

▶ **Theorem 21.** *Fix rational numbers $t_1 < t_2$. Suppose y_t has a zero in the interval (t_1, t_2) and y_t is not identically zero over this interval. There is k_0 as the smallest k such that z_t^k in* (18) *does* not *have any zero in (t_1, t_2). Moreover, the zero of y_t in (t_1, t_2) is non-tangential if k_0 is odd and is tangential if k_0 is even.*

Intuitively, the above theorem states that if y_t has at least one zero in (t_1, t_2), we can check for the existence of a tangential or non-tangential zero by a finite number of applications of Theorem 20 to functions z_t^k in (18). Note that y_t may have both tangential and non-tangential zeros; Theorem 21 gives a way of identifying the type of one of the zeros (the one with the largest order).

Proof of Theorem 21. Since y_t is an exponential polynomial, so is z_t^k for all k. Thus, we can use Theorem 20 to check if z_t^k has a zero in (t_1, t_2). Note that z_t^k is the sum of squares of $\frac{d^j}{dt^j}y_t$, which means

$$z_{t^*}^k = 0 \;\Rightarrow\; y_{t^*} = \frac{dy_t}{dt}\Big|_{t=t^*} = \cdots = \frac{d^k y_t}{dt^k}\Big|_{t=t^*} = 0. \tag{19}$$

The first part of the theorem is proved by showing that if for each k, z_t^k has a zero in (t_1, t_2), then y_t is identically zero. Suppose $z_t^k = 0$ for some $t = t_k^*$ in the interval (t_1, t_2), for any $k \in \{0,1,2,\ldots\}$. Using (19), we get that $y_t = 0$ for all $t \in \{t_0^*, t_1^*, t_2^*, \ldots\}$. If the set $\{t_0^*, t_1^*, t_2^*, \ldots\}$ is not finite, we get that y_t is identically zero according to the identity theorem [1]. If the set of zeros is finite, there is some t^* that appears infinitely often in the sequence $(t_0^*, t_1^*, t_2^*, \ldots)$. Therefore, $z_{t^*}^k = 0$ for infinitely many indices, which means $\frac{d^k y_t}{dt^k}\big|_{t=t^*} = 0$ for all k. Having y_k as an analytic function, this again implies that y_t is identically zero.

Since y_t is not identically zero, take k_0 such that $z_t^{k_0}$ does not have a zero in (t_1, t_2) but $z_t^{k_0-1}$ does. Then, there is $t^* \in (t_1, t_2)$ such that y_t and all its derivatives up to order $k_0 - 1$ are zero at t^* but $\frac{d^{k_0}}{dt^{k_0}}y_t\big|_{t=t^*} \neq 0$. This t^* and k_0 satisfy the conditions of Proposition 11. Thus, t^* is a non-tangential zero for y_t if k_0 is odd and a tangential zero if k_0 is even. ◀

To check if there is a non-tangential zero in an interval $(0, B)$, we apply Theorem 21 to each zero of y_t individually. Suppose y_t has at least one zero. We can localize all zeros of y_t as follows:

1. Set $(t_1, t_2) := (0, B)$;
2. Set k_0 to be the smallest index such that z_t^k in (18) does not have any zero in (t_1, t_2);
3. If $k_0 > 0$, do the next steps:
   - Use bisection to find an interval $(t', t'') \subset (t_1, t_2)$ such that over this interval, $z_t^{k_0-1}$ has a zero and $z_t^{k_0}$ and $\frac{d^{k_0}}{dt^{k_0}} y_t$ do not have any zero;
   - Store (t', t'');
   - Repeat Steps 2-3 with $(t_1, t_2) := (t_1, t')$;
   - Repeat Steps 2-3 with $(t_1, t_2) := (t'', t_2)$.

The bisection used in the above algorithm sequentially splits the interval into two sub-intervals and picks the one that contains the zero of $z_t^{k_0-1}$. It stops when $\frac{d^{k_0}}{dt^{k_0}} y_t$ does not have any zero over the selected sub-interval. The splitting terminates after a finite number of iterations due to the fact that $\frac{d^{k_0}}{dt^{k_0}} y_t$ is a continuous function and non-zero at the zero of y_t. The whole algorithm terminates after a finite number of iterations since y_t has a finite number of zeros in $(0, B)$ (note that if y_t has infinite number of zeros in $(0, B)$, it will be identically zero according to the identity theorem [1]). The output of the algorithm is a set of intervals. Within each interval, y_t has a single zero. Applying Theorem 21 to each such interval will decide whether the zero is tangential or non-tangential.

References

1 Mark J. Ablowitz and Athanassios S. Fokas. *Complex Variables: Introduction and Applications*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2003. doi:10.1017/CBO9780511791246.

2 S. Akshay, Timos Antonopoulos, Joël Ouaknine, and James Worrell. Reachability problems for Markov chains. *Inf. Process. Lett.*, 115(2):155–158, 2015. doi:10.1016/j.ipl.2014.08.013.

3 Ana Medina Ayala, Sean B. Andersson, and Calin Belta. Formal synthesis of control policies for continuous time Markov processes from time-bounded temporal logic specifications. *IEEE Trans. Automat. Contr.*, 59(9):2568–2573, 2014. doi:10.1109/TAC.2014.2309033.

4 Adnan Aziz, Kumud Sanwal, Vigyan Singhal, and Robert K. Brayton. Model-checking continous-time Markov chains. *ACM Trans. Comput. Log.*, 1(1):162–170, 2000. doi:10.1145/343369.343402.

5 Christel Baier, Holger Hermanns, Joost-Pieter Katoen, and Boudewijn R. Haverkort. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *Theor. Comput. Sci.*, 345(1):2–26, 2005. doi:10.1016/j.tcs.2005.07.022.

6 Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.

7 Paul C. Bell, Jean-Charles Delvenne, Raphaël M. Jungers, and Vincent D. Blondel. The continuous Skolem-Pisot problem. *Theor. Comput. Sci.*, 411(40-42):3625–3634, 2010. doi:10.1016/j.tcs.2010.06.005.

8 Tomás Brázdil, Vojtech Forejt, Jan Krcál, Jan Kretínský, and Antonín Kucera. Continuous-time stochastic games with time-bounded reachability. *Inf. Comput.*, 224:46–70, 2013. doi:10.1016/j.ic.2013.01.001.

9 Peter Buchholz, Ernst Moritz Hahn, Holger Hermanns, and Lijun Zhang. Model checking algorithms for CTMDPs. In *Computer Aided Verification - 23rd International Conference, CAV 2011, Snowbird, UT, USA, July 14-20, 2011. Proceedings*, volume 6806 of *Lecture Notes in Computer Science*, pages 225–242. Springer, 2011. doi:10.1007/978-3-642-22110-1_19.

10 Peter Buchholz and Ingo Schulz. Numerical analysis of continuous time Markov decision processes over finite horizons. *Comput. Oper. Res.*, 38(3):651–659, 2011. doi:10.1016/j.cor.2010.08.011.

11 Ventsislav Chonev, Joël Ouaknine, and James Worrell. On the Skolem problem for continuous linear dynamical systems. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 100:1–100:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.ICALP.2016.100.

12 Laure Daviaud, Marcin Jurdzinski, Ranko Lazic, Filip Mazowiecki, Guillermo A. Pérez, and James Worrell. When is containment decidable for probabilistic automata? In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, volume 107 of *LIPIcs*, pages 121:1–121:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ICALP.2018.121.

13 John Fearnley, Markus N. Rabe, Sven Schewe, and Lijun Zhang. Efficient approximation of optimal control for continuous-time Markov games. *Inf. Comput.*, 247:106–129, 2016. doi:10.1016/j.ic.2015.12.002.

14 Serge Lang. *Introduction to transcendental numbers.* Addison-Wesley series in mathematics. Addison-Wesley Pub. Co., 1966.

15 Serge Lang. Transcendental numbers and Diophantine approximations. *Bull. Amer. Math. Soc.*, 77(5):635–677, September 1971.

16 Angus Macintyre. Model theory of exponentials on Lie algebras. *Math. Struct. Comput. Sci.*, 18(1):189–204, 2008. doi:10.1017/S0960129508006622.

17 Angus Macintyre. Turing meets Schanuel. *Ann. Pure Appl. Log.*, 167(10):901–938, 2016. doi:10.1016/j.apal.2015.10.003.

18 Angus Macintyre and Alex J. Wilkie. On the decidability of the real exponential field. In *Kreiseliana. About and Around Georg Kreisel*, pages 441–467. A K Peters, 1996.

19 Bruce L. Miller. Finite state continuous time Markov decision processes with a finite planning horizon. *SIAM Journal on Control*, 6(2):266–280, 1968.

20 Martin R. Neuhäußer, Mariëlle Stoelinga, and Joost-Pieter Katoen. Delayed nondeterminism in continuous-time Markov decision processes. In *Foundations of Software Science and Computational Structures, 12th International Conference, FOSSACS 2009, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2009, York, UK, March 22-29, 2009. Proceedings*, volume 5504 of *Lecture Notes in Computer Science*, pages 364–379. Springer, 2009. doi:10.1007/978-3-642-00596-1_26.

21 Martin R. Neuhäußer and Lijun Zhang. Time-bounded reachability probabilities in continuous-time Markov decision processes. In *QEST 2010, Seventh International Conference on the Quantitative Evaluation of Systems, Williamsburg, Virginia, USA, 15-18 September 2010*, pages 209–218. IEEE Computer Society, 2010. doi:10.1109/QEST.2010.47.

22 Martin R. Neuhäußer. *Model checking nondeterministic and randomly timed systems.* PhD thesis, Univ. Twente, Enschede, 2010.

23 Jakob Piribauer and Christel Baier. On Skolem-hardness and saturation points in Markov decision processes. In *ICALP*, 2020.

24 Markus N. Rabe and Sven Schewe. Finite optimal control for time-bounded reachability in CTMDPs and continuous-time Markov games. *Acta Inf.*, 48(5-6):291–315, 2011. doi:10.1007/s00236-011-0140-0.

25 Markus N. Rabe and Sven Schewe. Optimal time-abstract schedulers for CTMDPs and continuous-time Markov games. *Theor. Comput. Sci.*, 467:53–67, 2013. doi:10.1016/j.tcs.2012.10.001.

26 Mahmoud Salamati, Sadegh Soudjani, and Rupak Majumdar. A Lyapunov approach for time-bounded reachability of CTMCs and CTMDPs. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS)*, 5(1):1–29, 2020.

27 A. J. Wilkie. Schanuel's conjecture and the decidability of the real exponential field. In *Algebraic Model Theory*, pages 223–230. Springer Netherlands, Dordrecht, 1997.

28 Nicolás Wolovick and Sven Johr. A characterization of meaningful schedulers for continuous-time Markov decision processes. In *Formal Modeling and Analysis of Timed Systems, 4th International Conference, FORMATS 2006, Paris, France, September 25-27, 2006, Proceedings*, volume 4202 of *Lecture Notes in Computer Science*, pages 352–367. Springer, 2006. doi:10.1007/11867340_25.