# FI MU

**Faculty of Informatics**
**Masaryk University Brno**

# Continuous-Time Stochastic Games with Time-Bounded Reachability

**by**

**Tomáš Brázdil**
**Vojtěch Forejt**
**Jan Krčál**
**Jan Křetínský**
**Antonín Kučera**

**FI MU Report Series** **FIMU-RS-2009-09**

 **October 2009**

**Publications in the FI MU Report Series are in general accessible via WWW:**

`http://www.fi.muni.cz/reports/`

**Further information can be obtained by contacting:**

**Faculty of Informatics**
**Masaryk University**
**Botanická 68a**
**602 00 Brno**
**Czech Republic**

# Continuous-Time Stochastic Games with Time-Bounded Reachability*

Tomáš Brázdil Vojtěch Forejt Jan Krčál Jan Křetínský
Antonín Kučera
Faculty of Informatics, Masaryk University,
Botanická 68a, 60200 Brno,
Czech Republic
{brazdil,forejt,krcal,xkretins,kucera}@fi.muni.cz

**Abstract**

We study continuous-time stochastic games with time-bounded reachability objectives. We show that each vertex in such a game has a *value* (i.e., an equilibrium probability), and we classify the conditions under which optimal strategies exist. Finally, we show how to compute optimal strategies in finite uniform games, and how to compute $\varepsilon$-optimal strategies in finitely-branching games with bounded rates (for finite games, we provide detailed complexity estimations).

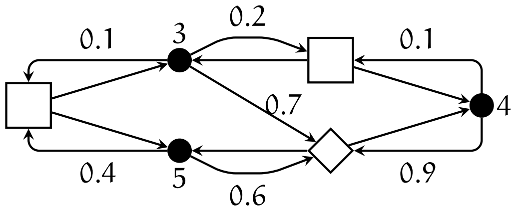
## 1 Introduction

Markov models are widely used in many diverse areas such as economics, biology, or physics. More recently, they have also been used for performance and dependability analysis of computer systems. Since faithful modeling of computer systems often requires both *randomized* and *non-deterministic* choice, a lot of attention has been devoted to Markov models where these two phenomena co-exist, such as *Markov decision processes* and *stochastic games*. The latter model of stochastic games is particularly apt for analyzing the interaction between a system and its environment, which are formalized as two *players* with antagonistic objectives (we refer to, e.g., [11, 6, 12] for more comprehensive expositions of results related to games in formal analysis and verification of

*Supported by Research Center "Institute for Theoretical Computer Science (ITI)", project No. 1M0545.

computer systems). So far, most of the existing results concern *discrete-time* Markov decision processes and stochastic games, and the accompanying theory is relatively well-developed (see, e.g., [10, 5]).

In this paper, we study *continuous-time stochastic games (CTGs)* and hence also *continuous-time Markov decision processes (CTMDPs)* with time-bounded reachability objectives. Roughly speaking, a CTG is a finite or countably infinite graph with three types of vertices—controllable vertices (boxes), adversarial vertices (diamonds), and actions (circles). The outgoing edges of controllable and adversarial vertices lead to the actions that are *enabled* at a given vertex. The outgoing edges of actions lead to controllable or adversarial vertices, and every edge is assigned a positive probability so that the total sum of these probabilities is equal to 1. Further, each action is assigned a positive real *rate*. A simple finite CTG is shown below.




A game is played by two players, $\Box$ and $\Diamond$, who are responsible for selecting the actions (i.e., resolving the non-deterministic choice) in the controllable and adversarial vertices, respectively. The selection is timeless, but performing a selected action takes time which is exponentially distributed (the parameter is the rate of a given action). When a given action is finished, the next vertex is chosen randomly according to the fixed probability distribution over the outgoing edges of the action. A *time-bounded reachability objective* is specified by a set $T$ of target vertices and a time bound $t > 0$. The goal of player $\Box$ is to maximize the probability of reaching a target vertex before time $t$, while player $\Diamond$ aims at minimizing this probability.

Note that events such as component failures, user requests, message receipts, exceptions, etc., are essentially history-independent, which means that the time between two successive occurrences of such events is exponentially distributed. CTGs provide a natural and convenient formal model for systems exhibiting these features, and time-bounded reachability objectives allow to formalize basic liveness and safety properties of these systems.

**Previous work.** Although the practical relevance of CTGs with time-bounded reachability objectives to verification problems is obvious, to the best of our knowledge there are no previous results concerning even very basic properties of such games. A more restricted model of uniform CTMDPs is studied in [4, 8]. Intuitively, a uniform CTMDP is a CTG where all non-deterministic vertices are controlled just by one player, and all actions are assigned the same rate. In [4], it is shown that the maximal and minimal probability of reaching a target vertex before time $t$ is efficiently computable up to an arbitrarily small given error, and that the associated strategy is also effectively computable. An open question explicitly raised in [4] is whether this result can be extended to all (not necessarily uniform) CTMDP. In [4], it is also shown that time-dependent strategies are more powerful than time-abstract ones, and this issue is addressed in greater detail in [8] where the mutual relationship between various classes of time-dependent strategies in CTMDPs is studied. Furthermore, in [3] reward-bounded objectives in CTMDPs are studied.

**Our contribution** is twofold. Firstly, we examine the *fundamental properties* of CTGs, where we aim at obtaining as general (and tight) results as possible. Secondly, we consider the associated *algorithmic issues*. Concrete results are discussed in the following paragraphs.

*Fundamental properties of CTGs.* We start by showing that each vertex $\hat{v}$ in a CTG with time-bounded reachability objectives has a *value*, i.e., an *equilibrium probability* of reaching a target vertex before time $t$. The value is equal to $\sup_\sigma \inf_\pi \mathcal{P}_{\hat{v}}^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$ and $\inf_\pi \sup_\sigma \mathcal{P}_{\hat{v}}^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$, where $\sigma$ and $\pi$ range over all time-abstract strategies of player $\Box$ and player $\Diamond$, and $\mathcal{P}_{\hat{v}}^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$ is the probability of reaching $T$ before time $t$ in a play obtained by applying the strategies $\sigma$ and $\pi$. This result holds for *arbitrary* CTGs which may have countably many vertices and actions. This immediately raises the question whether each player has an *optimal* strategy which achieves the outcome equal to or better than the value against every strategy of the opponent. We show that the answer is negative in general, but an optimal strategy for player $\Diamond$ is guaranteed to exist in *finitely-branching* CTGs, and an optimal strategy for player $\Box$ is guaranteed to exist in *finitely-branching* CTGs with *bounded rates* (see Definition 2.2). These results are tight, which is documented by appropriate counterexamples. Moreover, we show that in the subclasses of CTGs just mentioned, the players have also optimal CD strategies (a strategy is CD if it is deterministic and "counting", i.e., it only depends on the number of actions in the history of a play, where actions with the same rate are identified). Note that CD strategies still use infinite memory and in general they do not admit a finite description. A special attention is devoted to finite uniform CTGs, where we show a somewhat surprising result—both players have *finite memory optimal strategies* (these finite memory strategies are deterministic and their decision is based on "bounded counting" of actions; hence, we call them "BCD").

*Algorithms.* We show that for finite CTGs, $\varepsilon$-optimal strategies for both players are computable in $|V|^2 \cdot |A| \cdot bp \cdot (|\mathcal{R}|+1)^{\mathcal{O}((\max \mathcal{R})\cdot t + \ln \frac{1}{\varepsilon})}$ time, where $|V|$ and $|A|$ is the number of vertices and actions, resp., $bp$ is the maximum bit-length of transition probabilities and rates (we assume that rates and the probabilities in distributions assigned to the actions are represented as fractions of integers encoded in binary), $|\mathcal{R}|$ is the number of rates, $\max \mathcal{R}$ is the maximal rate, and $t$ is the time bound. This solves the open problem of [4] (in fact, our result is more general as it applies to finite CTGs, not just to finite CTMDPs). Actually, the algorithm works also for *infinite-state* CTGs as long as they are finitely-branching, have bounded rates, and satisfy some natural "effectivity assumptions" (see Corollary 4.2). For example, this is applicable to the class of infinite-state CTGs definable by pushdown automata (where the rate of a given configuration depends just on the current control state), and also to other automata-theoretic models. Finally, we show how to compute the optimal BCD strategies for both players in finite uniform CTGs.

All proofs have been moved into the appendix. In the main body of the paper, we just try to indicate basic ideas behind the proofs. This is not always possible, because some arguments are tricky and hard to explain at the intuitive level (occasionally we also rely on relatively advanced calculations). Nevertheless, the results themselves should be easy to understand.

# 2 Definitions

In this paper, the sets of all positive integers, non-negative integers, rational numbers, real numbers, non-negative real numbers, and positive real numbers are denoted by $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{R}^{\geq 0}$, and $\mathbb{R}^{>0}$, respectively. Let $A$ be a finite or countably infinite set. A *probability distribution* on $A$ is a function $f : A \to \mathbb{R}^{\geq 0}$ such that $\sum_{a \in A} f(a) = 1$. The *support* of $f$ is the set of all $a \in A$ where $f(a) > 0$. A distribution $f$ is *rational* if $f(a) \in \mathbb{Q}$ for every $a \in A$, *positive* if $f(a) > 0$ for every $a \in A$, and *Dirac* if $f(a) = 1$ for some $a \in A$. The set of all distributions on $A$ is denoted by $\mathcal{D}(A)$. A *$\sigma$-field* over a set $\Omega$ is a set $\mathcal{F} \subseteq 2^{\Omega}$ that includes $\Omega$ and is closed under complement and countable union. A *measurable space* is a pair $(\Omega, \mathcal{F})$ where $\Omega$ is a set called *sample space* and $\mathcal{F}$ is a $\sigma$-field over $\Omega$ whose elements are called *measurable sets*. A *probability measure* over a measurable space $(\Omega, \mathcal{F})$

is a function $\mathcal{P} : \mathcal{F} \to \mathbb{R}^{\geq 0}$ such that, for each countable collection $\{X_i\}_{i \in I}$ of pairwise disjoint elements of $\mathcal{F}$, $\mathcal{P}(\bigcup_{i \in I} X_i) = \sum_{i \in I} \mathcal{P}(X_i)$, and moreover $\mathcal{P}(\Omega) = 1$. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathcal{P})$, where $(\Omega, \mathcal{F})$ is a measurable space and $\mathcal{P}$ is a probability measure over $(\Omega, \mathcal{F})$. Given two measurable sets $X, Y \in \mathcal{F}$ such that $\mathcal{P}(Y) > 0$, the *conditional probability* of $X$ under the condition $Y$ is defined as $\mathcal{P}(X \mid Y) = \mathcal{P}(X \cap Y)/\mathcal{P}(Y)$. We say that a property $A \subseteq \Omega$ holds *for almost all* elements of a measurable set $Y$ if $\mathcal{P}(Y) > 0$, $A \cap Y \in \mathcal{F}$, and $\mathcal{P}((A \cap Y) \mid Y) = 1$.

In our next definition we introduce continuous-time Markov chains (CTMCs). The literature offers several equivalent definitions of CTMCs (see, e.g., [9]). For purposes of this paper, we adopt the variant where transitions have discrete probabilities and the rates are assigned to states.

**Definition 2.1.** *A* continuous-time Markov chain (CTMC) *is a tuple* $\mathcal{M} = (M, \rightarrow, \mathit{Prob}, \mathbf{R}, \mathit{Init})$, *where* $M$ *is a finite or countably infinite set of* states, $\rightarrow\ \subseteq M \times M$ *is a* transition relation *such that every* $s \in M$ *has at least one outgoing transition, Prob is a function which to each* $s \in M$ *assigns a positive probability distribution over the set of its outgoing transitions,* $\mathbf{R}$ *is a function which to each* $s \in M$ *assigns a positive real* rate, *and Init is the* initial probability distribution *on* $M$.

We write $s \xrightarrow{x} s'$ to indicate that $s \rightarrow s'$ and $\mathit{Prob}(s)(s \rightarrow s') = x$. A *time-abstract path* is a finite or infinite sequence $u = u_0, u_1, \ldots$ of states such that $u_{i-1} \rightarrow u_i$ for every $1 \leq i < \mathit{length}(u)$, where $\mathit{length}(u)$ is the length of $u$ (the length of an infinite sequence is $\infty$). A *timed path* (or just *path*) is a pair $w = (u, t)$, where $u$ is a time-abstract path and $t = t_1, t_2, \ldots$ is a sequence of positive reals such that $\mathit{length}(t) = \mathit{length}(u)$. We put $\mathit{length}(w) = \mathit{length}(u)$, and for every $0 \leq i < \mathit{length}(w)$, we usually write $w(i)$ and $w[i]$ instead of $u_i$ and $t_i$, respectively.

Infinite paths are also called *runs*. The set of all runs in $\mathcal{M}$ is denoted $\mathit{Run}_{\mathcal{M}}$, or just *Run* when $\mathcal{M}$ is clear from the context. A *template* is a pair $(u, I)$, where $u = u_0, u_1, \ldots$ is a finite time-abstract path and $I = I_0, I_1, \ldots$ a finite sequence of non-empty intervals in $\mathbb{R}^{\geq 0}$ such that $\mathit{length}(u) = \mathit{length}(I)$. Every template $(u, I)$ determines a *basic cylinder* $\mathit{Run}(u, I)$ consisting of all runs $w$ such that $w(i) = u_i$ for all $0 \leq i < \mathit{length}(u)$, and $w[j] \in I_j$ for all $0 \leq i < \mathit{length}(u) - 1$. To $\mathcal{M}$ we associate the probability space $(\mathit{Run}, \mathcal{F}, \mathcal{P})$ where $\mathcal{F}$ is the $\sigma$-field generated by all basic cylinders $\mathit{Run}(u, I)$ and $\mathcal{P} : \mathcal{F} \to \mathbb{R}^{\geq 0}$ is the unique probability measure on $\mathcal{F}$ such that

$$\mathcal{P}(\mathit{Run}(u, I)) = \mathit{Init}(u_0) \cdot \prod_{i=0}^{\mathit{length}(u)-2} \mathit{Prob}(u_i)(u_i \rightarrow u_{i+1}) \cdot \left(e^{-\mathbf{R}(u_i) \cdot \sup(I_i)} - e^{-\mathbf{R}(u_i) \cdot \inf(I_i)}\right)$$

Note that if $length(u) = 1$, the "big product" above is empty and hence equal to 1.

Now we formally define continuous-time games, which generalize continuous-time Markov chains by allowing not only probabilistic but also *non-deterministic* choice. Continuous-time games also generalize the model of continuous-time Markov decision processes studied in [4, 8] by splitting the non-deterministic vertices into two disjoint subsets of *controllable* and *adversarial* vertices, which are controlled by two "players" with antagonistic objectives. Thus, one can model the interaction between a system and its environment.

**Definition 2.2.** *A* continuous-time game (CTG) *is a tuple* $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$ *where* $V$ *is a finite or countably infinite set of* vertices, $A$ *is a finite or countably infinite set of* actions, $\mathbf{E}$ *is a function which to every* $v \in V$ *assigns a non-empty set of actions* enabled *in* $v$, $(V_\Box, V_\Diamond)$ *is a partition of* $V$, $\mathbf{P}$ *is a function which assigns to every* $a \in A$ *a probability distribution on* $V$, *and* $\mathbf{R}$ *is a function which assigns a positive real* rate *to every* $a \in A$.

*We require that* $V \cap A = \emptyset$ *and use* $N$ *to denote the set* $V \cup A$. *We say that* $G$ *is* finitely-branching *if for each* $v \in V$ *the set* $\mathbf{E}(v)$ *is finite (note that* $\mathbf{P}(a)$ *for a given* $a \in A$ *can still have an infinite support.) We say that* $G$ *has* bounded rates *if* $\sup_{a \in A} \mathbf{R}(a) < \infty$, *and that* $G$ *is* uniform *if* $\mathbf{R}$ *is a constant function. Finally, we say that* $G$ *is* finite *if both* $V$ *and* $A$ *are finite.*

If $V_\Box$ or $V_\Diamond$ is empty (i.e., there is just one type of vertices), then $G$ is a *continuous-time Markov decision process (CTMDP)*. Technically, our definition of CTMDP is slightly different from the one used in [4, 8], but the difference is only cosmetic. The two models are equivalent in a well-defined sense (a detailed explanation is included in Appendix B). Also note that $\mathbf{P}$ and $\mathbf{R}$ associate the probability distributions and rates directly to actions, not to pairs of $V \times A$. This is perhaps somewhat non-standard, but leads to simpler notation (since each vertex can have its "private" set of enabled actions, this is no restriction).

A *play* of $G$ is initiated in some vertex. The non-deterministic choice is resolved by two players, $\Box$ and $\Diamond$, who select the actions in the vertices of $V_\Box$ and $V_\Diamond$, respectively. The selection itself is timeless, but some time is spent by performing the selected action (the time is exponentially distributed with the rate $\mathbf{R}(a)$), and then a transition to the next vertex is chosen randomly according to the distribution $\mathbf{P}(a)$. The players can also select the actions *randomly*, i.e., they select not just a single action but a *probability distribution* on the enabled actions. Moreover, the players are allowed to play differently when the same vertex is revisited. We assume that both players can see the history of a play, but cannot measure the elapsed time.

Let $\odot \in \{\Box, \Diamond\}$. A *strategy* for player $\odot$ is a function which to each $wv \in \mathbb{N}^* V_\odot$ assigns a probability distribution on $\mathbf{E}(v)$. The sets of all strategies for player $\Box$ and player $\Diamond$ are denoted by $\Sigma$ and $\Pi$, respectively. Each pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ together with an initial vertex $\hat{v} \in V$ determine a unique *play* of the game G, which is a CTMC $G(\hat{v}, \sigma, \pi)$ where $\mathbb{N}^* A$ is the set of states, the rate of a given $wa \in \mathbb{N}^* A$ is $\mathbf{R}(a)$ (the rate function of $G(\hat{v}, \sigma, \pi)$ is also denoted by $\mathbf{R}$), and transitions exist only between states of the form $wa$ and $wava'$, where $wa \xrightarrow{x} wava'$ iff one of the following conditions is satisfied:

- $v \in V_\Box$, $a' \in \mathbf{E}(v)$, and $x = \mathbf{P}(a)(v) \cdot \sigma(wv)(a') > 0$
- $v \in V_\Diamond$, $a' \in \mathbf{E}(v)$, and $x = \mathbf{P}(a)(v) \cdot \pi(wv)(a') > 0$

The initial distribution is determined as follows:

- $Init(\hat{v}a) = \sigma(\hat{v})(a)$ if $\hat{v} \in V_\Box$ and $a \in \mathbf{E}(\hat{v})$;
- $Init(\hat{v}a) = \pi(\hat{v})(a)$ if $\hat{v} \in V_\Diamond$ and $a \in \mathbf{E}(\hat{v})$;
- in the other cases, *Init* returns zero.

Note that the set of states of $G(\hat{v}, \sigma, \pi)$ is infinite. Also note that all states reachable from a state $\hat{v}a$, where $Init(\hat{v}a) > 0$, are alternating sequences of vertices and actions. We say that a state $w$ of $G(\hat{v}, \sigma, \pi)$ *hits* a vertex $v \in V$ if $v$ is the last vertex which appears in $w$ (for example, $v_1 a_1 v_2 a_2$ hits $v_2$). Further, we say that $w$ hits $T \subseteq V$ if $w$ hits some vertex of T. From now on, the paths (both finite and infinite) in $G(\hat{v}, \sigma, \pi)$ are denoted by Greek letters $\alpha, \beta, \ldots$. Note that for every $\alpha \in Run_{G(\hat{v},\sigma,\pi)}$ and every $i \in \mathbb{N}_0$ we have that $\alpha(i) = wa$ where $wa \in \mathbb{N}^* A$.

We denote by $\mathcal{R}(G)$ the set of all rates used in G (i.e., $\mathcal{R}(G) = \{\mathbf{R}(a) \mid a \in A\}$), and by $\mathcal{H}(G)$ the set of all vectors of the form $\mathbf{i} : \mathcal{R}(G) \to \mathbb{N}_0$ satisfying $\sum_{r \in \mathcal{R}(G)} \mathbf{i}(r) < \infty$. When G is clear from the context, we write just $\mathcal{R}$ and $\mathcal{H}$ instead of $\mathcal{R}(G)$ and $\mathcal{H}(G)$, respectively. For every $\mathbf{i} \in \mathcal{H}$, we put $|\mathbf{i}| = \sum_{r \in \mathcal{R}} \mathbf{i}(r)$. For every $r \in \mathcal{R}$, we denote by $\mathbf{1}_r$ the vector of $\mathcal{H}$ such that $\mathbf{1}_r(r) = 1$ and $\mathbf{1}_r(r') = 0$ if $r' \neq r$. Further, for every $wx \in \mathbb{N}^* \mathbb{N}$ we define the vector $\mathbf{i}_{wx} \in \mathcal{H}$ such that $\mathbf{i}_{wx}(r)$ returns the cardinality of the set $\{j \in \mathbb{N}_0 \mid 0 \leq j < length(w), w(j) \in A, \mathbf{R}(w(j)) = r\}$ (Note that the last element $x$ of $wx$ is disregarded.) Given $\mathbf{i} \in \mathcal{H}$ and $wx \in \mathbb{N}^* \mathbb{N}$, we say that $wx$ *matches* $\mathbf{i}$ if $\mathbf{i} = \mathbf{i}_{wx}$.

We say that a strategy $\tau$ is *counting (C)* if $\tau(wv) = \tau(w'v)$ for all $w, w' \in \mathbb{N}^*$ such that $\mathbf{i}_{wv} = \mathbf{i}_{w'v}$. In other words, a strategy $\tau$ is counting if the only information about the history of a play $w$ which influences the decision of $\tau$ is the vector $\mathbf{i}_{wv}$. Hence, every counting strategy $\tau$ can be considered as a function from $\mathcal{H} \times V$ to $\mathcal{D}(A)$, where $\tau(\mathbf{i}, v)$ corresponds to the value of $\tau(wv)$ for every $wv$ matching $\mathbf{i}$. A counting strategy $\tau$ is *bounded counting (BC)* if there is $k \in \mathbb{N}$ such that $\tau(wv) = \tau(w'v)$ whenever $|w|, |w'| \geq k$. A strategy $\tau$ is *deterministic (D)* if $\tau(wv)$ is a Dirac distribution for all $wv$. Strategies that are not necessarily counting are called *history-dependent (H)*, and strategies that are not necessarily deterministic are called *randomized (R)*. Thus, we obtain the following six types of strategies: BCD, BCR, CD, CR, HD, and HR. The most general (unrestricted) type is HR, and the importance of the other types of strategies becomes clear in subsequent sections.

In this paper, we are interested in continuous-time games with *time-bounded reachability objectives*, which are specified by a set $T \subseteq V$ of *target* vertices and a *time bound* $t \in \mathbb{R}^{>0}$. The goal of player $\Box$ is to maximize the probability of reaching a target vertex before the time bound t, while player $\Diamond$ aims at minimizing this probability. Let $\hat{v}$ be the initial vertex. Then each pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ determines a unique *outcome* $\mathcal{P}^{\sigma,\pi}_{\hat{v}}(Reach^{\leq t}(T))$, which is the probability of all $\alpha \in Run_{G(\hat{v},\sigma,\pi)}$ that visit T before time t (i.e., there is $i \in \mathbb{N}_0$ such that $\alpha(i)$ hits T and $\sum_{i=0}^{i-1} \alpha[i] \leq t$). A fundamental question (answered in Section 3) is whether continuous-time games with time-bounded reachability objectives have a *value*, i.e., a unique equilibrium outcome. We say that $\hat{v} \in V$ *has a value* if

$$\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(Reach^{\leq t}(T)) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(Reach^{\leq t}(T))$$

If $\hat{v}$ has a value, then $val(\hat{v})$ denotes the *value of* $\hat{v}$ defined by the above equality. Further, if $\hat{v}$ has a value, it makes sense to define *ε-optimal* and *optimal* strategies in $\hat{v}$. Let $\varepsilon \geq 0$. We say that a strategy $\sigma \in \Sigma$ is an *ε-optimal maximizing strategy in* $\hat{v}$ (or just *ε-optimal in* $\hat{v}$) if

$$\inf_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(Reach^{\leq t}(T)) \;\geq\; val(\hat{v}) - \varepsilon\,,$$

and that a strategy $\pi \in \Pi$ is an *ε-optimal minimizing strategy in* $\hat{v}$ (or just *ε-optimal in* $\hat{v}$) if

$$\sup_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(Reach^{\leq t}(T)) \;\leq\; val(v) + \varepsilon$$

A strategy is *optimal* in $\hat{v}$ if it is 0-optimal in $\hat{v}$, and just *optimal* if it is optimal in every $\hat{v}$.

# 3 The Existence of Values and Optimal Strategies

In this section we first prove that every vertex in a CTG with time-bounded reachability objectives has a value. This result holds without any additional restrictions (i.e., for

CTGs with possibly countable state-space and infinite branching degree). From this we immediately obtain the existence of $\varepsilon$-optimal strategies for both players for every $\varepsilon > 0$. Then, we study the existence of optimal strategies. We show that even though optimal minimizing strategies may not exist in infinitely-branching CTGs, they always exist in finitely-branching ones. As for optimal maximizing strategies, we show that they do not necessarily exist even in finitely-branching CTGs, but they are guaranteed to exist if a game is both finitely-branching and has bounded rates (see Definition 2.2).

For the rest of this section, we fix a CTG $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$, a set $T \subseteq V$ of target vertices, and a time bound $t > 0$. Given $\mathbf{i} \in \mathcal{H}$ where $|\mathbf{i}| > 0$, we denote by $F_{\mathbf{i}}$ the probability distribution function of the random variable $\sum_{r\in\mathcal{R}} \sum_{i=1}^{\mathbf{i}(r)} X_i^{(r)}$ where all $X_i^{(r)}$ are mutually independent and each $X_i^{(r)}$ is an exponentially distributed random variable with the rate $r$ (for reader's convenience, basic properties of exponentially distributed random variables are recalled in Appendix A). We also define $F_{\mathbf{0}}$ as a constant function returning 1 for every argument (here $\mathbf{0} \in \mathcal{H}$ is the empty history, i.e., $|\mathbf{0}| = 0$). In the special case when $\mathcal{R}$ is a singleton, we use $F_\ell$ and $f_\ell$ to denote $F_{\mathbf{i}}$ and $f_{\mathbf{i}}$ such that $\mathbf{i}(r) = \ell$, where $r$ is the only element of $\mathcal{R}$. Further, given $\sim \in \{<, \leq, =\}$ and $k \in \mathbb{N}$, we denote by $\mathcal{P}_{\hat{v}}^{\sigma,\pi}(\mathit{Reach}_{\sim k}^{\leq t}(T))$ the probability of all $\alpha \in \mathit{Run}_{G(\hat{v},\sigma,\pi)}$ that visit $T$ for the first time in the number of steps satisfying $\sim k$ and before time $t$ (i.e., there is $i \in \mathbb{N}_0$ such that $i = \min\{j \mid \alpha(j) \text{ hits } T\} \sim k$ and $\sum_{i=0}^{i-1} \alpha[i] \leq t$).

The following theorem says that every vertex in a CTG with bounded reachability objectives has a value. Let us note that the powerful result of Martin [7] about weak determinacy of Blackwell games cannot be applied in this setting, at least not immediately. As we shall see, the ideas presented in the proof of Theorem 3.1 are useful also for designing an algorithm which for a given $\varepsilon > 0$ computes $\varepsilon$-optimal strategies for both players.

**Theorem 3.1.** *Every vertex $v \in V$ has a value.*

Roughly speaking, Theorem 3.1 is proved in the following way. Given $\sigma \in \Sigma$, $\pi \in \Pi$, $\mathbf{j} \in \mathcal{H}$, and $u \in V$, we denote by $P^{\sigma,\pi}(u, \mathbf{j})$ the probability of all runs $\alpha \in \mathit{Run}_{G(u,\sigma,\pi)}$ such that for some $n \in \mathbb{N}_0$ the state $\alpha(n)$ hits $T$ and matches $\mathbf{j}$, and for all $0 \leq j < n$ we have that $\alpha(j)$ does not hit $T$. Then we introduce two functions $\mathcal{A}, \mathcal{B} : \mathcal{H} \times V \to [0, 1]$ where

$$\mathcal{A}(\mathbf{i}, v) = \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \sum_{\mathbf{j}\in\mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j}) \qquad \mathcal{B}(\mathbf{i}, v) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \sum_{\mathbf{j}\in\mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j})$$

Intuitively, $\mathcal{A}(\mathbf{i}, v)$ and $\mathcal{B}(\mathbf{i}, v)$ give the "best" probability achievable by player $\Box$ and player $\Diamond$ in a vertex $v$, assuming that the history of a play matches $\mathbf{i}$. Hence, it suffices to prove that $\mathcal{A} = \mathcal{B}$, because then also $\mathcal{A}(\mathbf{0}, v) = \mathcal{B}(\mathbf{0}, v) = val(v)$, where $\mathbf{0}$ returns zero for every argument. The equality $\mathcal{A} = \mathcal{B}$ is obtained by demonstrating that both $\mathcal{A}$ and $\mathcal{B}$ are equal to the (unique) least fixed point of a monotonic function $\mathcal{V} : (\mathcal{H} \times V \to [0,1]) \to (\mathcal{H} \times V \to [0,1])$ defined as follows: for every $H : \mathcal{H} \times V \to [0,1]$, $\mathbf{i} \in \mathcal{H}$, and $v \in V$ we have that

$$\mathcal{V}(H)(\mathbf{i}, v) = \begin{cases} F_{\mathbf{i}}(t) & v \in T \\ \sup_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot H(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & v \in V_\Box \setminus T \\ \inf_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot H(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & v \in V_\Diamond \setminus T \end{cases}$$

The details are technical and can be found in Appendix C.

Observe that due to Theorem 3.1, both players have $\varepsilon$-optimal strategies in every vertex $v$ (for every $\varepsilon > 0$). This follows directly from the definition of $val(v)$ given in Section 2. Now we examine the existence of *optimal* strategies. We start by observing that optimal minimizing and optimal maximizing strategies do not necessarily exist, even if we restrict ourselves to games with finitely many rates (i.e., $\mathcal{R}(G)$ is finite) and finitely many distinct transition probabilities.

**Observation 3.2.** *Optimal minimizing and optimal maximizing strategies in continuous-time games with time-bounded reachability objectives do not necessarily exist, even if we restrict ourselves to games with finitely many rates (i.e., $\mathcal{R}(G)$ is finite) and finitely many distinct transition probabilities.*

However, if G is finitely-branching, then the existence of an optimal minimizing CD strategy can be established by adapting the construction used in the proof of Theorem 3.1. Observe that we do not require that G has bounded rates.

**Theorem 3.3.** *If G is finitely-branching, then there is an optimal minimizing CD strategy.*

The issue with optimal maximizing strategies is slightly more complicated. First, we observe that optimal maximizing strategies do not necessarily exist even in finitely-branching games.

**Observation 3.4.** *Optimal maximizing strategies in continuous-time games with time-bounded reachability objectives may not exist, even if we restrict ourselves to finitely-branching games.*

Now we show that if G is finitely-branching *and* has bounded rates, then there is an optimal maximizing CD strategy. To achieve that, we introduce the notion of k-*step optimal* strategies, which optimize the outcome in finite plays of length k. Observe that, due to Theorem 3.1, for all $k \in \mathbb{N}$ and $v \in V$ we have that $\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq k}^{\leq t}(T)) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq k}^{\leq t}(T))$. We use $\mathit{val}^k(v)$ to denote the $k$-*step value* defined by this equality, and we say that strategies $\sigma^k \in \Sigma$ and $\pi^k \in \Pi$ are $k$-*step optimal* if for all $v \in V$, $\pi \in \Pi$, and $\sigma \in \Sigma$ we have $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma^k,\pi}(\mathit{Reach}_{\leq k}^{\leq t}(T)) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi^k}(\mathit{Reach}_{\leq k}^{\leq t}(T)) = \mathit{val}^k(v)$. The existence and basic properties of $k$-step optimal strategies are stated in our next lemma.

**Lemma 3.5.** *If* $G$ *is finitely-branching and has bounded rates, then we have the following:*

1. *For all* $\varepsilon > 0$, $k \geq (\sup \mathcal{R})te^2 - \ln \varepsilon$, $\sigma \in \Sigma$, $\pi \in \Pi$, *and* $v \in V$ *we have that*

$$\mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T)) - \varepsilon \;\leq\; \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq k}^{\leq t}(T)) \;\leq\; \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$$

2. *For every* $k \in \mathbb{N}$, *there are* $k$*-step optimal BCD strategies* $\sigma^k \in \Sigma$ *and* $\pi^k \in \Pi$. *Further, for all* $\varepsilon > 0$ *and* $k \geq (\sup \mathcal{R})te^2 - \ln \varepsilon$ *we have that every* $k$*-step optimal strategy is also an* $\varepsilon$*-optimal strategy.*

If $G$ is finitely-branching and has bounded rates, one may be tempted to construct an optimal maximizing strategy $\sigma$ by selecting those actions that are selected by infinitely many $k$-step optimal BCD strategies for all $k \in \mathbb{N}$ (these strategies are guaranteed to exist by Lemma 3.5 (2)). However, this is not so straightforward, because the distributions assigned to actions in finitely-branching games can still have an infinite support. Intuitively, this issue is overcome by considering larger and larger finite subsets of the support so that the total probability of all of the infinitely many omitted elements approaches zero. Hence, a proof of the following theorem is somewhat technical.

**Theorem 3.6.** *If* $G$ *is finitely-branching and has bounded rates, then there is an optimal maximizing CD strategy.*

## 3.1 Optimal Strategies in Finite Uniform CTGs

In this subsection, we restrict ourselves to finite uniform CTGs and prove that both players have *optimal BCD strategies* in such games. Roughly speaking, the result is obtained by showing that optimal CD strategies (which are guaranteed to exist by Theorem 3.3 and Theorem 3.6) can be safely redefined into *greedy* strategies after performing a finite (and effectively computable) number of steps. Greedy strategies try to maximize/minimize the probability of reaching $T$ in as few steps as possible, and hence they can ignore the history of a play. Hence, the original optimal CD strategies become stationary after a finite number of steps, which means that they are in fact BCD. We also show that this result is tight in the sense that optimal BCD strategies do not necessarily exist in uniform CTGs with infinitely many states. In Section 4, we use these results to design an algorithm which *computes* the optimal BCD strategies in finite uniform games.

In this subsection, we assume that the previously fixed CTG $G$ is finite and that $\mathbf{R}(a) = r > 0$ for all $a \in A$. We start by introducing greedy strategies.

**Definition 3.7.** *A strategy* $\sigma \in \Sigma$ *is* greedily maximizing *if for all* $v \in V$ *and* $\sigma' \in \Sigma$ *one of the following two conditions is satisfied:*

- *For all* $i \in \mathbb{N}_0$ *we have* $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq i}^{<\infty}(T)) = \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma',\pi}(\mathit{Reach}_{\leq i}^{<\infty}(T))$.
- *There is* $i \in \mathbb{N}_0$ *such that* $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq i}^{<\infty}(T)) > \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma',\pi}(\mathit{Reach}_{\leq i}^{<\infty}(T))$ *and for all* $j < i$ *we have* $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq j}^{<\infty}(T)) = \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma',\pi}(\mathit{Reach}_{\leq j}^{<\infty}(T))$.

*Similarly,* $\pi \in \Pi$ *is* greedily minimizing *if for all* $v \in V$ *and* $\pi' \in \Pi$ *one of the following conditions holds:*

- *For all* $i \in \mathbb{N}_0$ *we have* $\sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq i}^{<\infty}(T)) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi'}(\mathit{Reach}_{\leq i}^{<\infty}(T))$.
- *There is* $i \in \mathbb{N}_0$ *such that* $\sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq i}^{<\infty}(T)) < \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi'}(\mathit{Reach}_{\leq i}^{<\infty}(T))$ *and for all* $j < i$ *we have* $\sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\leq j}^{<\infty}(T)) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi'}(\mathit{Reach}_{\leq j}^{<\infty}(T))$.

*A strategy* $\tau$ *is* stationary *if* $\tau$ *is deterministic and* $\tau(wv)$ *depends just on* $v$ *for every vertex* $v$.

Note that time plays no role in greedily maximizing/minimizing strategies. Our next lemma reveals that greedy *stationary* strategies exist and can be effectively computed in polynomial time in finite CTGs.

**Lemma 3.8.** *There is a greedily maximizing stationary strategy* $\sigma_g$*, and a greedily minimizing stationary strategy* $\pi_g$*. Moreover, the strategies* $\sigma_g$ *and* $\pi_g$ *are computable in polynomial time.*

Now we can state the main theorem of this subsection.

**Theorem 3.9.** *Let* $\sigma_g$ *be a greedily maximizing stationary strategy, and* $\pi_g$ *a greedily minimizing stationary strategy. Let* $\sigma$ *be an optimal maximizing CD strategy, and* $\pi$ *an optimal minimizing CD strategy. Then for all sufficiently large* $k \in \mathbb{N}$ *we have that BCD strategies* $\sigma' \in \Sigma$ *and* $\pi' \in \Pi$ *defined by*

$$\sigma'(i,v) = \begin{cases} \sigma(i,v) & \text{if } i < k; \\ \sigma_g(v) & \text{otherwise.} \end{cases} \qquad \pi'(i,v) = \begin{cases} \pi(i,v) & \text{if } i < k; \\ \pi_g(v) & \text{otherwise.} \end{cases}$$

*are optimal. Moreover, if all transition probabilities in* $G$ *are rational, then* $\sigma'$ *and* $\pi'$ *are optimal for all* $k \geq rt(1 + m^{|A|^2 \cdot |V|^2})$*, where* $m$ *is the maximal denominator of transition probabilities.*

A natural question is whether Theorem 3.9 can be extended to infinite-state uniform CTGs. The question is answered in our next observation.

**Observation 3.10.** *Optimal BCD strategies do not necessarily exist in uniform infinite-state CTGs, even if they are finitely-branching and use only finitely many distinct transition probabilities.*

# 4 Algorithms

Now we present algorithms which compute $\varepsilon$-optimal BCD strategies in finitely-branching CTGs with bounded rates and optimal BCD strategies in finite uniform CTGs. In this section, we assume that all rates and distributions used in the considered CTGs are *rational*.

## 4.1 Computing $\varepsilon$-optimal BCD strategies

For the rest of this subsection, let us fix a CTG $G = (V, A, \mathbf{E}, (V_\square, V_\Diamond), \mathbf{P}, \mathbf{R})$, a set $T \subseteq V$ of target vertices, a time bound $t > 0$, and some $\varepsilon > 0$. For simplicity, let us first assume that $G$ is finite; as we shall see, our algorithm does not really depend on this assumption, as long as the game is finitely-branching, has bounded rates, and its structure can be effectively generated (see Corollary 4.2). Let $k = (\max \mathcal{R})te^2 - \ln(\frac{\varepsilon}{2})$. Then, due to Lemma 3.5, all $k$-step optimal strategies are $\frac{\varepsilon}{2}$-optimal.

For every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and for every $v \in V$, our algorithm computes an action $C(\mathbf{i}, v) \in \mathbf{E}(v)$ which represents the choice of the constructed $\varepsilon$-optimal BCD strategies $\sigma_\varepsilon \in \Sigma$ and $\pi_\varepsilon \in \Pi$. That is, for every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and for every $v \in V_\square$, we put $\sigma_\varepsilon(\mathbf{i}, v)(C(\mathbf{i}, v)) = 1$, and for the other arguments we define $\sigma_\varepsilon$ arbitrarily so that $\sigma_\varepsilon$ remains a BCD strategy. The strategy $\pi_\varepsilon$ is induced by the function $C$ in the same way.

To compute $C(\mathbf{i}, v)$, our algorithm uses a family of probabilities $R(\mathbf{i}, u)$ of reaching $T$ from $u$ before time $t$ in at most $k - |\mathbf{i}|$ steps using the strategies $\sigma_\varepsilon$ and $\pi_\varepsilon$ and assuming that the history matches $\mathbf{i}$. Actually, our algorithm computes the probabilities $R(\mathbf{i}, u)$ only up to a sufficiently small error so that the actions chosen by $C$ are "sufficiently optimal" (i.e., the strategies $\sigma_\varepsilon$ and $\pi_\varepsilon$ are $\varepsilon$-optimal, but they are not necessarily $k$-step optimal for the $k$ chosen above). Our algorithm works in two phases:

1. For $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, compute a number $\ell_{\mathbf{i}}(t) > 0$ such that $\frac{|F_{\mathbf{i}}(t) - \ell_{\mathbf{i}}(t)|}{F_{\mathbf{i}}(t)} \leq \frac{\varepsilon^{2|\mathbf{i}|+1}}{2^{2|\mathbf{i}|+1}}$. For every $a \in A$ and $u \in V$, compute a floating point representation $\mathbf{p}(a)(u)$ of $\mathbf{P}(a)(u)$ satisfying $\frac{|\mathbf{P}(a)(u) - \mathbf{p}(a)(u)|}{\mathbf{P}(a)(u)} \leq \frac{\varepsilon^{2k+1}}{2^{2k+1}}$.

2. Compute (in a bottom up fashion) the functions R and C as follows: Given $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$, we have that

$$R(\mathbf{i}, v) = \begin{cases} \ell_{\mathbf{i}}(t) & \text{if } v \in T \\ 0 & \text{if } v \notin T \text{ and } |\mathbf{i}| = k \\ \max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_{\Box} \setminus T \text{ and } |\mathbf{i}| < k \\ \min_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_{\Diamond} \setminus T \text{ and } |\mathbf{i}| < k \end{cases}$$

   For all $|\mathbf{i}| < k$ and $v \notin T$, we put $C(\mathbf{i}, v) = a$ where $a$ is an action that realizes the maximum (or minimum).

In Appendix D.1 we show that the strategies $\sigma_\varepsilon$ and $\pi_\varepsilon$ are indeed $\varepsilon$-optimal. Complexity analysis of the algorithm reveals the following (*bp* denotes the maximum bit-length of $\mathbf{P}(a)(v)$ and rates, assuming that we represent $\mathbf{P}(a)(v)$ and rates as fractions of integers encoded in binary).

**Theorem 4.1.** *Assume that* G *is finite. Then for every* $\varepsilon > 0$ *there are* $\varepsilon$*-optimal BCD strategies* $\sigma_\varepsilon \in \Sigma$ *and* $\pi_\varepsilon \in \Pi$ *computable in time* $|V|^2 \cdot |A| \cdot bp \cdot (|\mathcal{R}| + 1)^{\mathcal{O}((\max \mathcal{R}) \cdot t + \ln \frac{1}{\varepsilon})}$.

Note that our algorithm needs to analyze only a finite part of G. Hence, it also works for infinite games which satisfy the conditions formulated in the next corollary.

**Corollary 4.2.** *Let* G *be a finitely-branching game with bounded rates and let* $v \in V$*. Assume that the vertices and actions of* G *reachable from* $v$ *in a given finite number of steps are effectively computable, and that an upper bound on rates is also effectively computable. Then for every* $\varepsilon > 0$ *there are effectively computable BCD strategies* $\sigma_\varepsilon \in \Sigma$ *and* $\pi_\varepsilon \in \Pi$ *that are* $\varepsilon$*-optimal in* $v$*.*

## 4.2 Computing optimal BCD strategies in uniform finite games

For the rest of this subsection, we fix a finite uniform CTG $G = (V, A, \mathbf{E}, (V_{\Box}, V_{\Diamond}), \mathbf{P}, \mathbf{R})$ where $\mathbf{R}(a) = r > 0$ for all $a \in A$. Let $k = rt(1 + m^{|A|^2 \cdot |V|^2})$ (see Theorem 3.9).

The algorithm works similarly as the one of Section 4.1, but there are also some differences. Since we have just one rate, the vector $\mathbf{i}$ becomes just a number $i$. Similarly as in Section 4.1, our algorithm computes an action $C(i, v) \in \mathbf{E}(v)$ representing the choice of the constructed optimal BCD strategies $\sigma_{max} \in \Sigma$ and $\pi_{min} \in \Pi$. By Lemma 3.9, every optimal strategy can, from the k-th step on, start to behave as a fixed greedy stationary strategy, and we can compute such a greedy stationary strategy in polynomial time. Hence, the optimal BCD strategies $\sigma_{max}$ and $\pi_{min}$ are defined as follows:

$$\sigma_{max}(i, v) = \begin{cases} C(i, v) & \text{if } i < k; \\ \sigma_g(v) & \text{otherwise.} \end{cases} \qquad \pi_{min}(i, v) = \begin{cases} C(i, v) & \text{if } i < k; \\ \pi_g(v) & \text{otherwise.} \end{cases}$$

To compute the function C, our algorithm uses a table of symbolic representations of the (precise) probabilities $R(i, v)$ (here $i \leq k$ and $v \in V$) of reaching T from $v$ before time t in at most $k - i$ steps using the strategies $\sigma_{max}$ and $\pi_{min}$ and assuming that the history matches i.

The function C and the family of all $R(i, v)$ are computed (in a bottom up fashion) as follows: For all $0 \leq i \leq k$ and $v \in V$ we have that

$$R(i, v) = \begin{cases} F_i(t) & \text{if } v \in T \\ \sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma_g, \pi_g}(Reach_{=j}^{<\infty}(T)) & \text{if } v \notin T \text{ and } i = k \\ \max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i+1, u) & \text{if } v \in V_\Box \setminus T \text{ and } i < k \\ \min_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i+1, u) & \text{if } v \in V_\Diamond \setminus T \text{ and } i < k \end{cases}$$

For all $i < k$ and $v \in V$, we put $C(i, v) = a$ where $a$ is an action maximizing or minimizing $\sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i+1, u)$, depending on whether $v \in V_\Box$ or $v \in V_\Diamond$, respectively. The effectivity of computing such an action (this issue is not trivial) is discussed in the proof of the following theorem.

**Theorem 4.3.** *The BCD strategies $\sigma_{max}$ and $\pi_{min}$ are optimal and effectively computable.*

# 5 Conclusions, Future Work

We have shown that vertices in CTGs with time bounded reachability objectives have a value, and we classified the subclasses of CTGs where a given player has an optimal strategy. We also proved that in finite uniform CTGs, both players have optimal BCD strategies. Finally, we designed algorithms which compute $\varepsilon$-optimal BCD strategies in finitely-branching CTGs with bounded rates, and optimal BCD strategies in finite uniform CTGs.

There are several interesting directions for future research. First, we can consider more general classes of strategies that depend on the elapsed time (in our setting, strategies are time-abstract). In [4], it is demonstrated that time-dependent strategies can achieve better results than the time-abstract ones. Further, [8] shows the power of time-dependent strategies differs when the player knows only the time consumed by the last action, or the complete timed history of a play. It is not immediately clear whether Theorem 3.1 still holds for time-dependent strategies, and whether it makes sense to think about optimal strategies in this setting. Second, a generalization to semi-Markov processes and games, where arbitrary (not only exponential) distributions are considered, would be desirable. Another interesting open problem is the existence of optimal BCD strategies in (not necessarily uniform) games.

## References


[1] S.V. Amari and R.B. Misra. Closed-form expressions for distribution of sum of exponential random variables. *IEEE transactions on reliability*, 46:519–522, 1997.

[2] A. Aziz, K. Sanwal, V. Singhal, and R.K. Brayton. Model-checking continuous-time Markov chains. *ACM Trans. on Comp. Logic*, 1(1):162–170, 2000.

[3] C. Baier, B.R. Haverkort, H. Hermanns, and J.-P. Katoen. Reachability in continuous-time Markov reward decision processes. In E. Graedel, J. Flum, and T. Wilke, editors, *Logic and Automata: History and Perspectives*, volume 2 of Texts in Logics and Games, pages 53–Ű72. Amsterdam University Press, 2008.

[4] C. Baier, H. Hermanns, J.-P. Katoen, and B.R. Haverkort. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *TCS*, 345:2–26, 2005.

[5] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1996.

[6] E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics, and Infinite Games*. Number 2500 in LNCS. Springer, 2002.

[7] D.A. Martin. The determinacy of Blackwell games. *JSL*, 63(4):1565–1581, 1998.

[8] M. Neuhäußer, M. Stoelinga, and J.-P. Katoen. Delayed nondeterminism in continuous-time Markov decision processes. In *Proceedings of FoSSaCS 2009*, volume 5504 of *LNCS*, pages 364–379. Springer, 2009.

[9] J.R. Norris. *Markov Chains*. Cambridge University Press, 1998.

[10] M.L. Puterman. *Markov Decision Processes*. Wiley, 1994.

[11] W. Thomas. Infinite games and verification. In *Proceedings of CAV 2003*, volume 2725 of *LNCS*, pages 58–64. Springer, 2003.

[12] I. Walukiewicz. A landscape with games in the background. In *Proceedings of LICS 2004*, pages 356–366. IEEE, 2004.

# A Exponentially Distributed Random Variables

For reader's convenience, in this section we recall basic properties of exponentially distributed random variables.

A *random variable* over a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is a function $X : \Omega \to \mathbb{R}$ such that the set $\{\omega \in \Omega \mid X(\omega) \leq c\}$ is measurable for every $c \in \mathbb{R}$. We usually write just $X{\sim}c$ to denote the set $\{\omega \in \Omega \mid X(\omega) \sim c\}$, where $\sim$ is a comparison and $c \in \mathbb{R}$. The *expected value* of X is defined by the Lebesgue integral $\int_{\omega \in \Omega} X(\omega)\, d\mathcal{P}$. A function $f : \mathbb{R} \to \mathbb{R}^{\geq 0}$ is a *density* of a random variable X if for every $c \in \mathbb{R}$ we have that $\mathcal{P}(X{\leq}c) = \int_{-\infty}^{c} f(x)\, dx$. If a random variable X has a density function f, then the expected value of X can also be computed by a (Riemann) integral $\int_{-\infty}^{\infty} x \cdot f(x)\, dx$. Random variables X, Y are *independent* if for all $c, d \in \mathbb{R}$ we have that $\mathcal{P}(X{\leq}c \cap Y{\leq}d) = \mathcal{P}(X{\leq}c) \cdot \mathcal{P}(Y{\leq}d)$. If X and Y are independent random variables with density functions $f_X$ and $f_Y$, then the random variable $X + Y$ (defined by $X + Y(\omega) = X(\omega) + Y(\omega)$) has a density function f which is the *convolution* of $f_X$ and $f_Y$, i.e., $f(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x)\, dx$.

A random variable X has an *exponential distribution with rate* $\lambda$ if $\mathcal{P}(X \leq c) = 1 - e^{-\lambda c}$ for every $c \in \mathbb{R}^{\geq 0}$. The density function $f_X$ of X is then defined as $f_X(c) = \lambda e^{-\lambda c}$ for all $c \in \mathbb{R}^{\geq 0}$, and $f_X(c) = 0$ for all $c < 0$. The expected value of X is equal to $\int_{-\infty}^{\infty} x \cdot \lambda e^{-\lambda x} dx = 1/\lambda$.

**Lemma A.1.** *Let* $\mathcal{M} = (M, \to, \mathit{Prob}, \mathbf{R}, \mathit{Init})$ *be a CTMC,* $j \in \mathbb{N}_0$, $t \in \mathbb{R}^{\geq 0}$, *and* $u_0, \ldots, u_j \in M$. *Let* U *be the set of all runs* $(u, t)$ *where* u *starts with* $u_0, \ldots, u_j$ *and* $\sum_{i=0}^{j} t_j \leq t$ . *We have that*

$$\mathcal{P}(U) = F_{\mathbf{i}}(t) \cdot \mathit{Init}(u_0) \cdot \prod_{\ell=0}^{j-1} \mathit{Prob}(u_\ell \to u_{\ell+1})$$

*where* $\mathbf{i}$ *assigns to every rate* r *the cardinality of the set* $\{k \mid \mathbf{R}(u_k) = r, 0 \leq k \leq j\}$

PROOF. By induction on j. For $j = 0$ the lemma holds, because we $\mathcal{P}(U) = \mathit{Init}(u_0)$ by definition.

Now suppose that $j > 0$ and the lemma holds for all $k < j$. We denote by $U_k^{t'}$ the set of all runs $(u, t)$ where $u$ starts with $u_0, \ldots, u_j$ and $\sum_{i=0}^{k} t_i = t'$. We have that

$$
\begin{aligned}
\mathcal{P}(U) &= \int_0^t \mathcal{P}(U_{j-1}^x) \cdot \mathit{Prob}(u_{j-1} \to u_j) \cdot e^{-\mathbf{R}(u_{j-1}) \cdot (t-x)} \, dx \\
&= \int_0^t F_{\mathbf{i} - \mathbf{1}_{\mathbf{R}(u_{j-1})}}(x) \cdot \left( \prod_{\ell=0}^{j-2} \mathit{Prob}(u_\ell \to u_{\ell+1}) \right) \mathit{Prob}(u_{j-1} \to u_j) \cdot e^{-\mathbf{R}(u_{j-1}) \cdot (t-x)} \, dx \\
&= \prod_{\ell=0}^{j-1} \mathit{Prob}(u_\ell \to u_{\ell+1}) \cdot \int_0^t F_{\mathbf{i} - \mathbf{1}_{\mathbf{R}(u_{j-1})}}(x) \cdot e^{-\mathbf{R}(u_{j-1}) \cdot (t-x)} \, dx \\
&= F_{\mathbf{i}}(x) \cdot \prod_{\ell=0}^{j-1} \mathit{Prob}(u_\ell \to u_{\ell+1})
\end{aligned}
$$

□

# B A Comparison of the Existing Definitions of CTMDPs

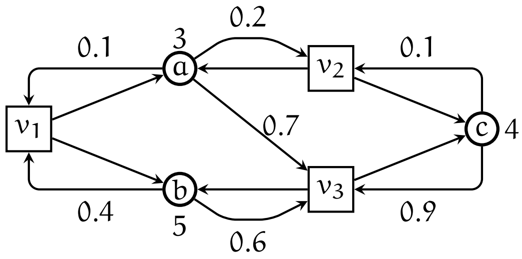
As we already mentioned in Section 2, our definition of CTG (and hence also CTMDP) is somewhat different from the definition of CTMDP used in [4, 8]. To prevent misunderstandings, we discuss the issue in greater detail in here and show that the two formalisms are in fact equivalent. First, let us recall the alternative definition CTMDP used in [4, 8].

**Definition B.1.** *A CTMDP is a triple $\mathcal{M} = (S, A, \mathbf{R})$, where $M$ is a finite or countably infinite set of* states, *$A$ is a finite or countably infinite set of* actions, *and $\mathbf{R} : (S \times A \times S) \to \mathbb{R}^{\geq 0}$ is a* rate matrix.

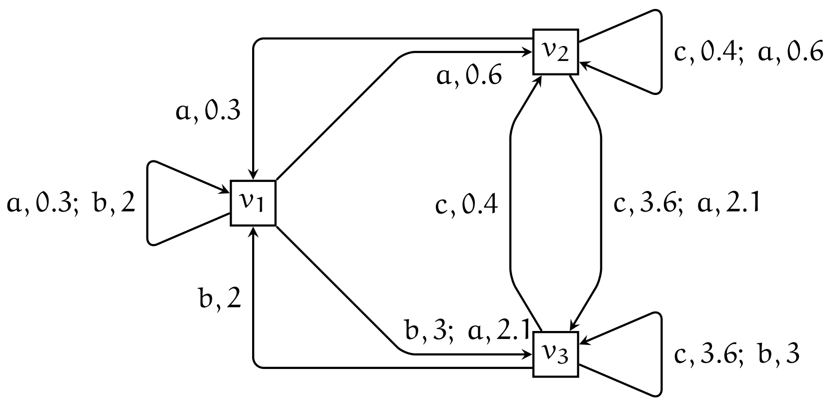
A CTMDP $\mathcal{M} = (S, A, \mathbf{R})$ can be depicted as a graph where $S$ is the set of vertices and $s \to s'$ is an edge labeled by $(a, r)$ iff $\mathbf{R}(s, a, s) = r > 0$. The conditional probability of selecting the edge $s \xrightarrow{(a,r)} s'$, under the condition that the action $a$ is used, is defined as $r/\mathbf{R}(s, a)$, where $\mathbf{R}(s, a) = \sum_{s \xrightarrow{(a,\hat{r})} \hat{s}} \hat{r}$. The time needed to perform the action $a$ in $s$ is exponentially distributed with the rate $\mathbf{R}(s, a)$. This means that $\mathcal{M}$ can be translated into an equivalent CTG where $S$ is the set of vertices, the set of actions is

$$\{(s, a) \mid s \in S, a \in A, \mathbf{R}(s, a, s') > 0 \text{ for some } s' \in S\}$$

where the rate of a given action $(s, a)$ is $\mathbf{R}(s, a)$, and $\mathbf{P}(s, a)(s') = \mathbf{R}(s, a, s')/\mathbf{R}(s, a)$. This translation also works in the opposite direction (assuming that $V = V_\Box$ or $V = V_\Diamond$). To illustrate this, consider the following CTG:




An equivalent CTMDP (in the sense of Definition B.1) looks as follows:




However, there is one subtle issue regarding strategies. In [4, 8], a strategy (controller) selects an action in every vertex. The selection may depend on the history of a play. In [4, 8], it is noted that if a controller is deterministic, then the resulting play is a CTMC. If a controller is *randomized*, one has to add "intermediate" discrete-time states which implement the timeless randomized choice, and hence the resulting play is *not* a CTMC, but a mixture of discrete-time and continuous-time Markov chains. In our setting, this problem disappears, because the probability distribution chosen by a player is simply "multiplied" with the probabilities of outgoing edges of actions. For deterministic strategies, the two approaches are of course completely equivalent.

# C Proofs of Section 3

## C.1 Proof of Theorem 3.1

**THEOREM 3.1.** *Every vertex $v \in V$ has a value.*

PROOF. Given $\sigma \in \Sigma$, $\pi \in \Pi$, $\mathbf{j} \in \mathcal{H}$, and $u \in V$, we denote by $P^{\sigma,\pi}(u, \mathbf{j})$ the probability of all runs $\alpha \in Run_{G(u,\sigma,\pi)}$ such that for some $n \in \mathbb{N}_0$ the state $\alpha(n)$ hits $T$ and matches $\mathbf{j}$, and for all $0 \leq j < n$ we have that $\alpha(j)$ does not hit $T$. Then we introduce two functions $\mathcal{A}, \mathcal{B} : \mathcal{H} \times V \to [0,1]$ where

$$\mathcal{A}(\mathbf{i}, v) = \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j}) \qquad \mathcal{B}(\mathbf{i}, v) = \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j})$$

Clearly, it suffices to prove that $\mathcal{A} = \mathcal{B}$, because then also $\mathcal{A}(\mathbf{0}, v) = \mathcal{B}(\mathbf{0}, v) = \mathit{val}(v)$, where $\mathbf{0}$ returns zero for every argument. The equality $\mathcal{A} = \mathcal{B}$ is obtained by demonstrating that both $\mathcal{A}$ and $\mathcal{B}$ are equal to the (unique) least fixed point of a monotonic function $\mathcal{V} : (\mathcal{H} \times V \to [0,1]) \to (\mathcal{H} \times V \to [0,1])$ defined as follows: for every $H : \mathcal{H} \times V \to [0,1]$, $\mathbf{i} \in \mathcal{H}$, and $v \in V$ we have that

$$\mathcal{V}(H)(\mathbf{i}, v) = \begin{cases} F_{\mathbf{i}}(t) & v \in T \\ \sup_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot H(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & v \in V_{\Box} \setminus T \\ \inf_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot H(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & v \in V_{\Diamond} \setminus T \end{cases}$$

Let us denote by $\mu\mathcal{V}$ the least fixed point of $\mathcal{V}$. We show that $\mu\mathcal{V} = \mathcal{A} = \mathcal{B}$. The inequality $\mathcal{A} \preceq \mathcal{B}$ is obvious and follows directly from the definition of $\mathcal{A}$ and $\mathcal{B}$. Hence, it suffices to prove the following two claims:

1. $\mathcal{A}$ is a fixed point of $\mathcal{V}$ (from this we obtain $\mu\mathcal{V} \preceq \mathcal{A}$).

2. For every $\varepsilon > 0$ there is a CD strategy $\pi_\varepsilon \in \Pi$ such that for every $\mathbf{i} \in \mathcal{H}$ and every $v \in V$ we have that

$$\sup_{\sigma \in \Sigma} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi_\varepsilon}(v, \mathbf{j}) \quad \leq \quad \mu(\mathcal{V})(\mathbf{i}, v) + \varepsilon$$

   (from this we get $\mathcal{B} \preceq \mu\mathcal{V}$).

**ad 1.** If $v \in T$, we have

$$\mathcal{A}(\mathbf{i}, v) = \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} F_{\mathbf{i}}(t) = \mathcal{V}(\mathcal{A})(\mathbf{i}, v)$$

Assume that $v \notin T$. Given a strategy $\tau \in \Sigma \cup \Pi$ and $a \in A$, we denote by $\tau^a$ a strategy defined by $\tau^a(wu) := \tau(vawu)$.

If $v \in V_\Box$,

$$\begin{aligned}
\mathcal{V}(A)(\mathbf{i}, v) &= \sup_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \sup_{d \in \mathcal{D}(\mathbf{E}(v))} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \sup_{d \in \mathcal{D}(\mathbf{E}(v))} \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \sup_{d \in \mathcal{D}(\mathbf{E}(v))} \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(a) > 0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(a) > 0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(a) > 0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \sum_{\substack{a \in A \\ \mathbf{j}(a) > 0}} \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) P^{\sigma,\pi}(u, \mathbf{j})
\end{aligned}$$

If $v \in V_\Diamond$,

$$\begin{aligned}
\mathcal{V}(A)(\mathbf{i}, v) &= \inf_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \inf_{d \in \mathcal{D}(\mathbf{E}(v))} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \inf_{d \in \mathcal{D}(\mathbf{E}(v))} \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \sup_{\sigma \in \Sigma} \inf_{d \in \mathcal{D}(\mathbf{E}(v))} \inf_{\pi \in \Pi} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(u, \mathbf{j}) \\
&= \sup_{\sigma \in \Sigma} \inf_{d \in \mathcal{D}(\mathbf{E}(v))} \inf_{\pi \in \Pi} \sum_{a \in A} d(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(a) > 0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(a) > 0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{a \in A} \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(a) > 0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \sum_{\substack{a \in A \\ \mathbf{j}(a) > 0}} \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot P^{\sigma^a,\pi^a}(u, \mathbf{j} - \mathbf{1}_{\mathbf{R}(a)}) \\
&= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \sum_{\mathbf{j} \in \mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) P^{\sigma,\pi}(u, \mathbf{j})
\end{aligned}$$

**ad 2.** The strategy $\pi_\varepsilon$ can be defined as follows. Given $\mathbf{i} \in \mathcal{H}$ and $v \in V_\Diamond$, we put $\pi_\varepsilon(\mathbf{i}, v)(a) = 1$ for some $a \in A$ satisfying $\sum_{u \in V} \mathbf{P}(a)(u) \cdot \mu\mathcal{V}(\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)}, u) \leq \mu\mathcal{V}(\mathbf{i}, v) + \frac{\varepsilon}{2^{|\mathbf{i}|+1}}$. We prove that $\pi_\varepsilon$ is $\varepsilon$-optimal minimizing. For every $\sigma \in \Sigma$, every $\mathbf{i} \in \mathcal{H}$, every $v \in V$ and every $k \geq 0$, we denote

$$\mathcal{R}^\sigma_k(\mathbf{i}, v) := \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ |\mathbf{j}| \leq k}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi_\varepsilon[\mathbf{i}]}(v, \mathbf{j})$$

Here $\pi_\varepsilon[\mathbf{i}]$ is the strategy obtained from $\pi_\varepsilon$ by $\pi_\varepsilon[\mathbf{i}](\mathbf{j}, u) := \pi_\varepsilon(\mathbf{i} + \mathbf{j}, u)$.

We prove that $\mathcal{R}^\sigma_k(\mathbf{i}, v) \leq \mu\mathcal{V}(\mathbf{i}, v) + \sum_{j=1}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}}$, which implies that $\mathcal{R}^\sigma(\mathbf{i}, v) = \lim_{k \to \infty} \mathcal{R}^\sigma_k(\mathbf{i}, v) \leq \mu\mathcal{V}(\mathbf{i}, v) + \varepsilon$.

For $v \in T$ we have

$$\mathcal{R}^\sigma_k(\mathbf{i}, v) = F_{\mathbf{i}}(t) = \mu\mathcal{V}(\mathbf{i}, v)$$

Assume that $v \notin T$. We proceed by induction on $k$. For $k = 0$ we have $\mathcal{R}_k^\sigma(\mathbf{i}, v) = 0 \leq \mu\mathcal{V}(\mathbf{i}, v)$. Assume that $v \in V_\Box \setminus T$

$$\begin{aligned}
\mathcal{R}_k^\sigma(\mathbf{i}, v) &= \sum_{a \in \mathbf{E}(v)} \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot \mathcal{R}_{k-1}^{\sigma^a}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&\leq \sum_{a \in \mathbf{E}(v)} \sigma(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot \left( \mu\mathcal{V}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) + \sum_{j=2}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}} \right) \\
&= \left( \sum_{a \in \mathbf{E}(v)} \sigma(v)(a) \cdot \sum_{u \in V} \mathbf{P}(a)(u) \cdot \mu\mathcal{V}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \right) + \sum_{j=2}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}} \\
&\leq \mu\mathcal{V}(\mathbf{i}, v) + \sum_{j=2}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}}
\end{aligned}$$

Finally, assume that $v \in V_\Diamond \setminus T$, and let $a \in A$ be the action such that $\pi_\varepsilon(\mathbf{i}, v)(a) = 1$

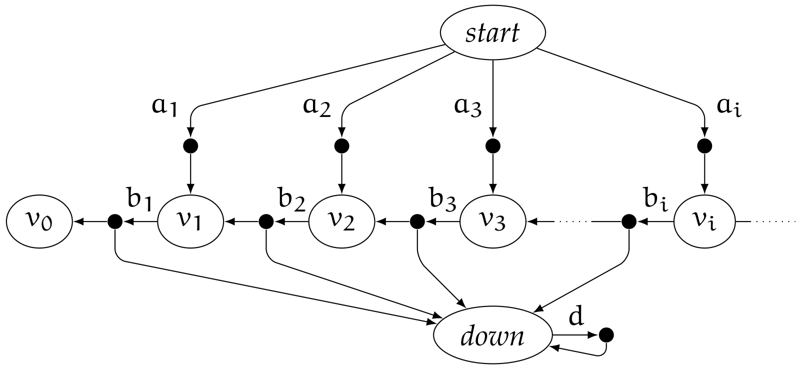
$$\begin{aligned}
\mathcal{R}_k^\sigma(\mathbf{i}, v) &= \sum_{u \in V} \mathbf{P}(a)(u) \cdot \mathcal{R}_{k-1}^{\sigma^a}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&\leq \sum_{u \in V} \mathbf{P}(a)(u) \cdot \left( \mu\mathcal{V}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) + \sum_{j=2}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}} \right) \\
&= \left( \sum_{u \in V} \mathbf{P}(a)(u) \cdot \mu\mathcal{V}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \right) + \sum_{j=2}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}} \\
&\leq \mu\mathcal{V}(\mathbf{i}, v) + \frac{\varepsilon}{2^{|\mathbf{i}|}} + \sum_{j=2}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}} \\
&\leq \mu\mathcal{V}(\mathbf{i}, v) + \sum_{j=1}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}}
\end{aligned}$$

□

## C.2 Proof of Observation 3.2

**OBSERVATION 3.2.** *Optimal minimizing and optimal maximizing strategies in continuous-time games with time-bounded reachability objectives do not necessarily exist, even if we restrict ourselves to games with finitely many rates (i.e., $\mathcal{R}(G)$ is finite) and finitely many distinct transition probabilities.*

PROOF. Consider a game $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$, where $V = \{v_i | i \in \mathbb{N}_0\} \cup \{\mathit{start}, \mathit{down}\}$, $A = \{a_i, b_i \mid i \in \mathbb{N}\} \cup \{d\}$, $\mathbf{E}(\mathit{start}) = \{a_i \mid i \in \mathbb{N}\}$, $\mathbf{E}(\mathit{down}) = \{d\}$, and $\mathbf{E}(v_i) = \{b_i\}$ for all $i \in \mathbb{N}$, $\mathbf{P}(a_i)(v_i) = 1$, $\mathbf{P}(d)(\mathit{down}) = 1$, and $\mathbf{P}(b_i)$ is the uniform distribution that chooses *down* and $v_{i-1}$ for all $i \in \mathbb{N}$, and $\mathbf{R}$ assigns 1 to every action. The structure of G is shown below (the partition of V into $(V_\square, V_\Diamond)$ is not fixed yet, and the vertices are therefore drawn as ovals).




If we put $V_\square = V$, we obtain that $\sup_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi}_{start}(Reach^{\leq 1}(\{down\})) = \sum_{\ell=1}^{\infty} \left(\frac{1}{2^\ell} F_\ell(1)\right)$ where $\pi$ is the trivial strategy for player $\Diamond$. However, there is obviously no optimal maximizing strategy. On the other hand, if we put $V_\Diamond = V$, we have that $\inf_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{start}(Reach^{\leq 1}(\{v_0\})) = 0$ where $\sigma$ is the trivial strategy for player $\square$, but there is no optimal minimizing strategy. $\square$

## C.3 Proof of Theorem 3.3

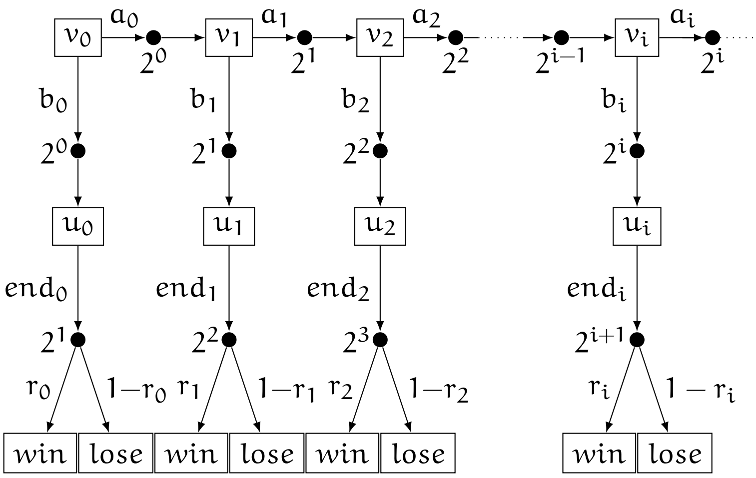
**THEOREM 3.3.** *If* G *is finitely-branching, then there is an optimal minimizing CD strategy.*

PROOF. It suffices to reconsider the proof of Claim 2 in the proof of Theorem 3.1, where G is finitely-branching and $\varepsilon$ is set to zero. Then the strategy $\pi_\varepsilon$ constructed in the proof of Claim 2 becomes an optimal minimizing CD strategy. $\square$

## C.4 Proof of Observation 3.4

**OBSERVATION 3.4.** *Optimal maximizing strategies in continuous-time games with time-bounded reachability objectives do not necessarily exist, even if we restrict ourselves to finitely-branching games.*

PROOF. Consider a game $G = (V, A, \mathbf{E}, (V_\square, V_\Diamond), \mathbf{P}, \mathbf{R})$, where $V = V_\square = \{v_i, u_i \mid i \in \mathbb{N}_0\} \cup \{win, lose\}$; $A = \{a_i, b_i, end_i \mid i \in \mathbb{N}_0\} \cup \{w, l\}$, $\mathbf{E}(win) = \{w\}$, $\mathbf{E}(lose) = \{l\}$, and $\mathbf{E}(v_i) = \{a_i, b_i\}$, $\mathbf{E}(u_i) = \{end_i\}$ for all $i \in \mathbb{N}_0$; $\mathbf{R}(win) = \mathbf{R}(lose) = 1$, and $\mathbf{R}(a_i) = \mathbf{R}(b_i) = 2^i$, $\mathbf{R}(end_i) = 2^{i+1}$ for all $i \in \mathbb{N}_0$; $\mathbf{P}(w)(win) = 1$, $\mathbf{P}(l)(lose) = 1$, and for all $i \in \mathbb{N}_0$ we have that $\mathbf{P}(a_i)(v_{i+1}) = 1$, $\mathbf{P}(b_i)(u_i) = 1$, and $\mathbf{P}(end_i)$ is the distribution that assigns $r_i$ to $win$ and $1 - r_i$ to *lose*, where $r_i$ is the number discussed below. The structure of G is shown below (note that for clarity, the vertices *win* and *lose* are drawn multiple times, and their only enabled actions $w$ and $l$ are not shown).




For every $k \in \mathbb{N}$, let $\mathbf{i}_k \in \mathcal{H}$ be the vector that assigns 1 to all $r \in \mathcal{R}$ such that $r \leq 2^k$, and 0 to all other rates. Let us fix $t \in \mathbb{Q}$ and $q \geq \frac{2}{3}$ such that $F_{\mathbf{i}_k}(t) \geq q$ for every $k \in \mathbb{N}$. This is indeed possible by Markov inequality and the mean of the variables' sum being less than 2. For every $j \geq 0$, we fix some $r_j \in \mathbb{Q}$ such that $q - \frac{1}{2^j} \leq F_{\mathbf{i}_{j+1}}(t) \cdot r_j \leq q - \frac{1}{2^{j+1}}$. It is easy to check that $r_j \in [0, 1]$, which means that the function $\mathbf{P}$ is well-defined.

We claim that $\sup_{\sigma \in \Sigma} \mathcal{P}_{v_0}^{\sigma,\pi}(\mathit{Reach}^{\leq t}(\{\mathit{win}\})) = q$ (where $\pi$ is the trivial strategy for player $\Diamond$), but there is no strategy $\sigma$ such that $\mathcal{P}_{s_0}^{\sigma}(\mathit{Reach}^{\leq t}(\{\mathit{win}\})) = q$. The first part follows by observing that player $\Box$ can reach *win* with probability at least $q - \frac{1}{2^j}$ for an arbitrarily large $j$ by selecting the actions $a_0, \ldots, a_{j-1}$ and then $b_j$. The second part follows from the fact that by using $b_j$, the probability of reaching *win* from $v_0$ becomes strictly lower than $q$, and by not selecting $b_j$ at all, this probability becomes equal to 0.

$\Box$

## C.5 Proof of Lemma 3.5

**LEMMA 3.5.** *Let us assume that* $G$ *is finitely-branching and has bounded rates. Then we have the following:*

1. *For every* $\varepsilon > 0$, $k \geq (\sup \mathcal{R})te^2 - \ln \varepsilon$, $\sigma \in \Sigma$, $\pi \in \Pi$, *and* $v \in V$ *we have that*

$$\mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T)) - \varepsilon \;\leq\; \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}_{\leq k}(T)) \;\leq\; \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$$

2. *For every* $k \in \mathbb{N}$, *there are* $k$*-step optimal BCD strategies* $\sigma^k \in \Sigma$ *and* $\pi^k \in \Pi$. *Further, for all* $\varepsilon > 0$ *and* $k \geq (\sup \mathcal{R})te^2 - \ln \varepsilon$ *we have that every* $k$*-step optimal strategy is also an* $\varepsilon$*-optimal strategy.*

PROOF. **ad 1.** Let us fix a rate $r = \sup \mathcal{R}$. It suffices to see that (here, the random variables used to define $F_i$ have rate $r$)

$$\begin{aligned}\sum_{n=k+1}^{\infty} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{=n}^{\leq t}(T)) &\leq \sum_{n=k+1}^{\infty} F_n(t) \cdot \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{=n}^{\leq \infty}(T)) \\ &\leq \sum_{n=k+1}^{\infty} F_{k+1}(t) \cdot \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{=n}^{\leq \infty}(T)) \\ &= F_{k+1}(t) \cdot \sum_{n=k+1}^{\infty} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{=n}^{\leq \infty}(T)) \\ &\leq F_{k+1}(t)\end{aligned}$$

which is less than $\varepsilon$ for $k \geq rte^2 - \ln \varepsilon$ by the following lemma.

**Lemma C.6.** *For every* $\varepsilon \in (0,1)$ *and* $n \geq rte^2 - \ln \varepsilon$ *we have* $F_n(t) < \varepsilon$.

PROOF.

$$F_n(t) = 1 - e^{-rt} \sum_{i=0}^{n-1} \frac{(rt)^i}{i!} = e^{-rt} \sum_{i=n}^{\infty} \frac{(rt)^i}{i!} = (*)$$

By Taylor's theorem for $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ and Lagrange form of the remainder we get

$$(*) \leq e^{-rt} \frac{(rt)^n}{n!} e^{rt} = \frac{(rt)^n}{n!} = (**)$$

By Stirling's formula $n! \approx \sqrt{n}(n/e)^n)$ we get

$$(**) < \left(\frac{rte}{n}\right)^n < \left(\frac{1}{e}\right)^{-\ln \varepsilon} = \varepsilon$$

by assumptions. □

**ad 2.** We proceed similarly as in the proof of Theorem 3.1 (we also use some notation of the proof of Theorem 3.1). Recall that given $\sigma \in \Sigma$, $\pi \in \Pi$, $\mathbf{j} \in \mathcal{H}$, and $u \in V$, we denote by $P^{\sigma,\pi}(u, \mathbf{j})$ the probability of all runs $\alpha \in \mathit{Run}_{G(u,\sigma,\pi)}$ such that for some $n \in \mathbb{N}_0$ the state $\alpha(n)$ hits $T$ and matches $\mathbf{j}$, and for all $0 \leq j < n$ we have that $\alpha(j)$ does not hit $T$.

Given $(\sigma, \pi) \in \Sigma \times \Pi$, $\mathbf{i} \in \mathcal{H}$ such that $|\mathbf{i}| \leq k$, and $v \in V$, we define

$$\bar{P}^{\sigma,\pi}(\mathbf{i}, v) := \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ |\mathbf{j}| \leq k - |\mathbf{i}|}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j})$$

the probability of reaching $T$ from $v$ before time $t$ in at most $k - |\mathbf{i}|$ steps using the strategies $\sigma$ and $\pi$ and assuming that the history matches $\mathbf{i}$.

To define the CD strategies $\sigma^k$ and $\pi^k$ we express the value $\sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \bar{P}^{\sigma,\pi}(\mathbf{i}, v)$ ($= \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \bar{P}^{\sigma,\pi}(\mathbf{i}, v)$, see below) using the following recurrence.

Given $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$, we define

$$\bar{R}(\mathbf{i}, v) := \begin{cases} F_{\mathbf{i}}(t) & \text{if } v \in T \\ 0 & \text{if } v \notin T \text{ and } |\mathbf{i}| = k \\ max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_{\Box} \setminus T \text{ and } |\mathbf{i}| < k \\ min_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_{\Diamond} \setminus T \text{ and } |\mathbf{i}| < k \end{cases}$$

For $v \notin T$ and $|\mathbf{i}| < k$ we define $\sigma^k(\mathbf{i}, v)$ and $\pi^k(\mathbf{i}, v)$ in the following way. If $v \in V_{\Box}$, we put $\sigma^k(\mathbf{i}, v)(a) = 1$ for some action $a$ which realizes the maximum in the definition of $\bar{R}(\mathbf{i}, v)$. Similarly, if $v \in V_{\Diamond}$, we put $\pi^k(\mathbf{i}, v)(a) = 1$ for some action $a$ which realizes the minimum in the definition of $\bar{R}(\mathbf{i}, v)$. For $|\mathbf{i}| \geq k$ and $v \in V$ we define $\sigma^k(\mathbf{i}, v)$ and $\pi^k(\mathbf{i}, v)$ arbitrarily so that $\sigma^k$ and $\pi^k$ remain BCD.

For every CD strategy $\tau \in \Sigma \cup \Pi$ and $\mathbf{i} \in \mathcal{H}$, we denote by $\tau[\mathbf{i}]$ the strategy obtained from $\tau$ by $\tau[\mathbf{i}](\mathbf{j}, u) := \tau(\mathbf{i} + \mathbf{j}, u)$.

Given $\pi \in \Pi$, $\mathbf{i} \in \mathcal{H}$ where $|\mathbf{i}| \leq k$, and $v \in V$, we define

$$Z^{\pi}(\mathbf{i}, v) := \bar{P}^{\sigma^k[\mathbf{i}], \pi}(\mathbf{i}, v)$$

Similarly, given $\sigma \in \Sigma$, $\mathbf{i} \in \mathcal{H}$ where $|\mathbf{i}| \leq k$, and $v \in V$, we define

$$Z^{\sigma}(\mathbf{i}, v) := \bar{P}^{\sigma, \pi^k[\mathbf{i}]}(\mathbf{i}, v)$$

We prove the following lemma.

**Lemma C.7.** *Let* $\mathbf{i} \in \mathcal{H}$, *where* $|\mathbf{i}| \leq k$, *and* $v \in V$. *Then*

$$\begin{aligned} \bar{R}(\mathbf{i}, v) &= \inf_{\pi \in \Pi} Z^{\pi}(\mathbf{i}, v) & (1) \\ &= \sup_{\sigma \in \Sigma} Z^{\sigma}(\mathbf{i}, v) & (2) \\ &= \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \bar{P}^{\sigma, \pi}(\mathbf{i}, v) & (3) \\ &= \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \bar{P}^{\sigma, \pi}(\mathbf{i}, v) & (4) \end{aligned}$$

*In particular, the strategies* $\sigma^k$ *and* $\pi^k$ *are* $k$*-step optimal because* $\bar{P}^{\sigma, \pi}(\mathbf{0}, v) = \mathcal{P}_v^{\sigma, \pi}(Reach_{\leq k}^{\leq t}(T))$.

PROOF. First, if $v \in T$, then for all $(\sigma, \pi) \in \Sigma \times \Pi$ we have $\bar{P}^{\sigma, \pi}(\mathbf{i}, v) = F_{\mathbf{i}}(t) = \bar{R}(\mathbf{i}, v)$. Assume that $v \notin T$. We proceed by induction on $n = k - |\mathbf{i}|$. For $n = 0$ we have $\bar{P}^{\sigma, \pi}(\mathbf{i}, v) = 0 = \bar{R}(\mathbf{i}, v)$. Assume the lemma holds for $n$, we show that it holds also for $n + 1$.

We start by proving the equation (1). Using the notation of the proof of Theorem 3.1, given a strategy $\tau \in \Sigma \cup \Pi$ and $a \in A$, we denote by $\tau^a$ a strategy defined by $\tau^a(wu) := \tau(vawu)$.

If $v \in V_\Box$ and $\sigma^k(\mathbf{i}, v)(b) = 1$,

$$\begin{aligned}
\inf_{\pi\in\Pi} Z^\pi(\mathbf{i}, v) &= \inf_{\pi\in\Pi} \sum_{u\in V} \mathbf{P}(b)(u) \cdot Z^{\pi^b}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \\
&= \sum_{u\in V} \mathbf{P}(b)(u) \cdot \inf_{\pi\in\Pi} Z^{\pi^b}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \\
&= \sum_{u\in V} \mathbf{P}(b)(u) \cdot \inf_{\pi\in\Pi} Z^{\pi}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \\
&= \sum_{u\in V} \mathbf{P}(b)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \\
&= \max_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&= \bar{R}(\mathbf{i}, v)
\end{aligned}$$

If $u \in V_\Diamond$,

$$\begin{aligned}
\inf_{\pi\in\Pi} Z^\pi(\mathbf{i}, v) &= \inf_{\pi\in\Pi} \sum_{a\in\mathbf{E}(v)} \pi(v)(a) \sum_{u\in V} \mathbf{P}(a)(u) \cdot Z^{\pi^a}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&= \inf_{d\in\mathcal{D}(\mathbf{E}(v))} \sum_{a\in\mathbf{E}(v)} d(a) \sum_{u\in V} \mathbf{P}(a)(u) \cdot \inf_{\pi\in\Pi} Z^{\pi^a}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&= \min_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot \inf_{\pi\in\Pi} Z^{\pi}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&= \min_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\
&= \bar{R}(\mathbf{i}, v)
\end{aligned}$$

The equation (2) can be proved in a similar manner.

The lemma follows from the following

$$\bar{R}(\mathbf{i}, v) = \inf_{\pi\in\Pi} Z^\pi(\mathbf{i}, v) \le \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \bar{P}^{\sigma,\pi}(\mathbf{i}, v) \le \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \bar{P}^{\sigma,\pi}(\mathbf{i}, v) \le \sup_{\sigma\in\Sigma} Z^\sigma(\mathbf{i}, v) = \bar{R}(\mathbf{i}, v)$$

□

The rest of the lemma is easily obtained from 1. as follows. Let $\varepsilon > 0$ and consider $k \geq (\sup \mathcal{R}) te^2 - \ln \varepsilon$. Then 1. implies that the value $\bar{R}(\mathbf{0}, v)$ of the k-step game initiated in $\nu$ satisfies $val(\nu) - \varepsilon \leq \bar{R}(\mathbf{0}, \nu) \leq val(\nu)$. Therefore all k-step optimal strategies are $\varepsilon$-optimal.

## C.6 Proof of Theorem 3.6

**THEOREM 3.6.** *If* $G$ *is finitely-branching and has bounded rates, then there is an optimal maximizing CD strategy.*

PROOF. For the sake of this proof, given a set of runs $R \subseteq Run_{G(\hat{\nu},\sigma,\pi)}$, we denote $\mathcal{P}^{\sigma,\pi}_{\hat{\nu}}(R)$ the probability of $R$ in $G(\hat{\nu}, \sigma, \pi)$. For every $k \in \mathbb{N}$ we fix a k-step optimal BCD strategy $\sigma_k$ of player $\Box$ (see Lemma 3.5). Let us order all numbers $r$ such that $\mathbf{R}(a) = r$ for some $a$ into an enumerable sequence $r_1, r_2 \ldots$ and the set $V$ into an enumerable sequence $\nu_1, \nu_2 \ldots$. We define a sequence of sets of strategies $\Sigma \supseteq \Gamma_0 \supseteq \Gamma_1 \supseteq \cdots$ as follows. We put $\Gamma_0 = \{\sigma_\ell \mid \ell \in \mathbb{N}\}$ and we construct $\Gamma_\ell$ to be an infinite subset of $\Gamma_{\ell-1}$ such that we have $\sigma(\mathbf{i}, \nu_n) = \sigma'(\mathbf{i}, \nu_n)$ for all $\sigma, \sigma' \in \Gamma_\ell$, all $n \leq \ell$ and all $\mathbf{i} \in \mathcal{H}$ such that $|\mathbf{i}| \leq \ell$ and $\mathbf{i}(r_j) = 0$ whenever $j > \ell$. Note that such a set exists since $\Gamma_{\ell-1}$ is infinite and the conditions above partition it into finitely many classes, one of which must be infinite.

Now we define the optimal strategy $\sigma$. Let $\mathbf{i} \in \mathcal{H}$ and $\nu_n \in V$, we choose a number $\ell$ such that $\ell > |\mathbf{i}|$, $\ell > n$ and $\mathbf{i}(j) = 0$ for all $j > \ell$ (note that such $\ell$ exists for any $\mathbf{i} \in \mathcal{H}$ and $\nu_n \in V$). We put $\sigma(\mathbf{i}, \nu_n) = \sigma'(\mathbf{i}, \nu_n)$ where $\sigma' \in \Gamma_\ell$. It is easy to see that $\sigma$ is a CD strategy, it remains to argue that it is optimal. Suppose the converse, i.e. that it is not $\varepsilon$-optimal in some $\nu_{in}$ for some $\varepsilon > 0$.

Let us fix $k$ satisfying conditions of part 1 of Lemma 3.5 for $\frac{\varepsilon}{4}$. For each $a \in A$ there is a set $B_a \subseteq V$ such that $V \setminus B_a$ is finite and $\mathbf{P}(a)(B_a) \leq \frac{\varepsilon}{4k}$. For all strategies $\sigma'$ and $\pi'$ and all $k$ we have that $\mathcal{P}^{\sigma',\pi'}_{\nu}(U^{\nu,\sigma',\pi'}_i) \leq \frac{\varepsilon}{2k}$ where $U^{\nu,\sigma',\pi'}_i$ is the set of all runs of $G(\nu, \sigma', \pi')$ that *do not* contain any state of the form $\nu_0 a_0 \ldots a_{i-1} \nu_{i-1} a_n$ where $\nu_{i-1} \in B_{a_{i-1}}$. As a consequence we have $\mathcal{P}^{\sigma',\pi'}_{\nu}(\bigcap_{i=1}^{k} U^{\nu,\sigma',\pi'}_i) \leq \frac{\varepsilon}{4}$. In the sequel, we denote $U^{\nu,\sigma',\pi'} = \bigcap_{i=1}^{k} U^{\nu,\sigma',\pi'}_i$ and we write just $U$ instead of $U^{\nu,\sigma',\pi'}$ if $\nu$, $\sigma$ and $\pi$ are clear from the context.

Let $W$ be the set of histories of the form $\nu_0 a_0 \ldots \nu_{i-1} a_{i-1} \nu_i$ where $i \leq k$, $\nu_0 = \nu_{in}$, and for all $0 \leq j < i$ we have $a_j \in \mathbf{E}(\nu_j)$, $\mathbf{P}(a_j)(\nu_{i+j}) > 0$ and $\nu_{j+1} \notin B_{a_j}$. We claim that there is $m \geq n$ s.t. $\sigma_m$ is $\frac{\varepsilon}{4}$-optimal and satisfies $\sigma(w) = \sigma_m(w)$ for all $w \in W$. To see that such a strategy exists, observe that $W$ is finite, which means that there is a number $\ell$ such that $k \leq \ell$ and for all $w \in W$, there is no $\nu_i$ in $w$ such that $i > \ell$ and whenever $a$ is in $w$, then $\mathbf{R}(a) = r_i$ for $i < \ell$. Now it suffices to choose arbitrary $\frac{\varepsilon}{4}$-optimal strategy $\sigma_m \in \Gamma_\ell$.

One can prove by induction on the length of path from $v_{in}$ to $T$ that the following equality holds true for all $\pi$.

$$\mathcal{P}^{\sigma_m,\pi}_{v_{in}}(Reach^{\leq t}_{\leq k}(T) \setminus U) = \mathcal{P}^{\sigma,\pi}_{v_{in}}(Reach^{\leq t}_{\leq k}(T) \setminus U)$$

Finally, we obtain

$$\begin{aligned}
\min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{v_{in}}(Reach^{\leq t}_{\leq k}(T) \setminus U) &= \min_{\pi\in\Pi} \mathcal{P}^{\sigma_m,\pi}_{v_{in}}(Reach^{\leq t}_{\leq k}(T) \setminus U) \\
&\geq \min_{\pi\in\Pi} \mathcal{P}^{\sigma_m,\pi}_{v_{in}}(Reach^{\leq t}_{\leq m}(T) \setminus U) - \frac{\varepsilon}{4} \\
&\geq \min_{\pi\in\Pi} \mathcal{P}^{\sigma_m,\pi}_{v_{in}}(Reach^{\leq t}_{\leq m}(T)) - \frac{\varepsilon}{2} \\
&\geq val(v_{in}) - \frac{\varepsilon}{4} - \frac{\varepsilon}{2} \geq val(v_{in}) - \varepsilon
\end{aligned}$$

which means that $\sigma$ is $\varepsilon$-optimal in $v_{in}$. □

## C.7 Proof of Lemma 3.8

**LEMMA 3.8.** *There is a greedily maximizing stationary strategy $\sigma_g$, and a greedily minimizing stationary strategy $\pi_g$. Moreover, the strategies $\sigma_g$ and $\pi_g$ are computable in polynomial time.*

PROOF. W.l.o.g. let us assume that all states in $T$ are absorbing, i.e. the only transitions leading from them are self-loops. Indeed, if the behavior of a greedy strategy is changed after reaching $T$ it still remains greedy. Also due to Theorem 3.3 and Theorem 3.6 we may restrict our attention to CD strategies. Therefore, in this proof, $\Sigma$ and $\Pi$ denote sets of CD strategies only.

The following algorithm computes which actions can be chosen in greedily minimizing and greedily maximizing strategies. We begin with the original game and keep on pruning inoptimal transitions until we reach a fix-point. In the first step, we compute the value $R_1(v)$ for each vertex $v$ which is the optimized probability of reaching $T$ in one step. We remove all transitions that are not optimal in this sense. In the next step, we consider reachability in precisely two steps. Note that we chose among the one-step optimal possibilities only. Transitions not optimal for two-steps reachability are removed and so forth. After stabilization, using the remaining transitions only thus results in greedy optimal behavior. Hence any such stationary strategy is a greedy optimal stationary strategy.

$$R_0(v) = \begin{cases} 1 & \text{if } v \in T, \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{E}_0(v) = \mathbf{E}(v)$$

$$R_{i+1}(a) = \sum_{u \in V} \mathit{Prob}(a)(u) \cdot R_i(u)$$

$$R_{i+1}(v) = \begin{cases} \max_{a \in \mathbf{E}_i(v)} R_{i+1}(a) & \text{if } v \in V_\Box, \\ \min_{a \in \mathbf{E}_i(v)} R_{i+1}(a) & \text{otherwise.} \end{cases}$$

$$\mathbf{E}_{i+1}(v) = \mathbf{E}_i(v) \cap \{a \mid R_{i+1}(a) = R_{i+1}(v)\}$$

In order to prove the correctness and state the complexity of the algorithm, we need the following definitions. We recall that a strategy $\tau[h]$ is defined by $\tau[h](j, u) = \tau(h + j, u)$. The probability to reach $T$ within time $t$ using strategies $\sigma$ and $\pi$ from $v$ having already taken $h$ steps *before* continuing from $v$ is denoted by

$$\mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{\leq t}(T)) = \sum_{j=0}^{\infty} F_{h+j}(t) \mathcal{P}^{\sigma[h],\pi[h]}_{v}(\mathit{Reach}^{<\infty}_{=i}(T))$$

Similarly, the probability to reach $T$ within $i$ steps using strategies $\sigma$ and $\pi$ from $v$ having already taken $h$ steps *before* continuing from $v$ is denoted by

$$\mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) = \mathcal{P}^{\sigma[h],\pi[h]}_{v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$$

Note that since there is no time limit here, it is determined by the underlying discrete Markov chain.

For $n \in \mathbb{N}$, we say that $\sigma \in \Sigma$ is *greedily maximizing on* $n$ *steps* if for every $h \in \mathbb{N}_0$ and $v \in V$, every $\sigma' \in \Sigma$ and every $i \leq n$ we have $\min_{\pi \in \Pi} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) = \min_{\pi \in \Pi} \mathcal{P}^{\sigma',\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$ unless there is $j \leq i$ satisfying $\min_{\pi \in \Pi} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq j}(T)) > \min_{\pi \in \Pi} \mathcal{P}^{\sigma',\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq j}(T))$. We write $\sigma \in \Sigma_{G(n)}$.

Similarly, for $n \in \mathbb{N}$, we say that $\sigma \in \Sigma$ is *greedily minimizing on* $n$ *steps* if for every $h \in \mathbb{N}_0$ and $v \in V$, every $\sigma' \in \Sigma$ and every $i \leq n$ we have $\max_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) = \max_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi'}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$ unless there is $j \leq i$ satisfying $\max_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq j}(T)) < \max_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi'}_{h,v}(\mathit{Reach}^{<\infty}_{\leq j}(T))$. We write $\pi \in \Pi_{G(n)}$.

A strategy $\tau \in \Sigma \cup \Pi$ is greedily optimizing if it is greedily maximizing or greedily minimizing.

**Lemma C.10.** *Strategies $\sigma \in \Sigma$ and $\pi \in \Pi$ are greedily maximizing on $n$ steps, and greedily minimizing on $n$ steps, respectively, iff they use transitions from $\mathbf{E}_n$ only.*

PROOF. $\Leftarrow$: We proceed by induction and prove that, moreover, for all $h \in \mathbb{N}_0$ and $v \in V$

$$\begin{aligned} R_n(v) &= \max_{\sigma\in\Sigma_{G(n-1)}} \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n}(T)) && = \min_{\pi\in\Pi} \mathcal{P}^{\sigma_n,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n}(T)) \\ &= \min_{\pi\in\Pi_{G(n-1)}} \max_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n}(T)) && = \max_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi_n}_{h,v}(\mathit{Reach}^{<\infty}_{\le n}(T)) \end{aligned}$$

for all strategies $\sigma_n, \pi_n$ using transitions from $\mathbf{E}_n$ only.

The case $n = 0$ is trivial. Now consider $n + 1$. We prove the maximizing part, the minimizing part is similar. Let $\sigma$ use transitions from $\mathbf{E}_{n+1}$ only. Using the induction hypothesis and the definition of greediness on $n + 1$ steps, it is sufficient to prove that $\min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n+1}(T)) \ge \min_{\pi\in\Pi} \mathcal{P}^{\sigma',\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n+1}(T))$ for all $h \in \mathbb{N}_0$ and $v \in V$ and $\sigma' \in \Sigma_{G(n)}$, i.e. those using transitions in $\mathbf{E}_n$ only. For $v \in V_\Box \setminus T$,

$$\begin{aligned} \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n+1}(T)) &= \min_{\pi\in\Pi} \sum_{u\in V} \sigma(h,v)(u)\mathcal{P}^{\sigma,\pi}_{h+1,u}(\mathit{Reach}^{<\infty}_{\le n}(T)) \\ &= \sum_{u\in V} \sigma(h,v)(u) \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{h+1,u}(\mathit{Reach}^{<\infty}_{\le n}(T)) \\ \text{(by IH and } \mathbf{E}_{n+1} \subseteq \mathbf{E}_n) &= \sum_{u\in V} \sigma(h,v)(u) \cdot R_n(u) \\ (\sigma \text{ uses } \mathbf{E}_{n+1} \text{ only}) &= \max_{a\in\mathbf{E}_n(v)} \sum_{u\in V} \mathit{Prob}(a)(u) \cdot R_n(u) = (*) \\ \text{(by IH)} &= \max_{a\in\mathbf{E}_n(v)} \sum_{u\in V} \mathit{Prob}(a)(u) \max_{\sigma\in\Sigma_{G(n-1)}} \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{h+1,u}(\mathit{Reach}^{<\infty}_{\le n}(T)) \\ &= \max_{\sigma\in\Sigma_{G(n)}} \min_{\pi\in\Pi} \mathcal{P}^{\sigma',\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\le n+1}(T)) \end{aligned}$$

Since $(*) = R_{n+1}(v)$, the equality with the first and the last expression proves the two auxiliary induction assumptions. For $v \in V_\Diamond \setminus T$, we simply use the induction hypothesis for the successors of $v$. And to prove the two auxiliary induction assumptions, we need to show that

$$R_{n+1}(v) = \min_{a\in\mathbf{E}_i(v)} \sum_{u\in V} \mathit{Prob}(a)(u) \cdot R_n(u) = \min_{a\in\mathbf{E}_i(v)} \sum_{u\in V} \mathit{Prob}(a)(u) \cdot \max_{\sigma\in\Sigma_{G(n-1)}} \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{u}(\mathit{Reach}^{<\infty}_{\le n}(T))$$

is (i) equal to

$$\max_{\sigma\in\Sigma_{G(n)}} \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{v}(\mathit{Reach}^{<\infty}_{\le n+1}(T)) = \max_{\sigma\in\Sigma_{G(n)}} \min_{\pi\in\Pi} \min_{a\in\mathbf{E}_i(v)} \sum_{u\in V} \mathit{Prob}(a)(u)\mathcal{P}^{\sigma,\pi}_{u}(\mathit{Reach}^{<\infty}_{\le n}(T))$$

which is clear and (ii) equal to

$$\min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{v}(Reach^{<\infty}_{\leq n+1}(\mathsf{T})) = \min_{a\in \mathbf{E}_i(v)} \sum_{u\in V} Prob(a)(u) \min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{u}(Reach^{<\infty}_{\leq n}(\mathsf{T}))$$

for every $\sigma$ using transitions from $\mathbf{E}_{n+1}$ only, which follows by IH, since $\min_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{u}(Reach^{<\infty}_{\leq n}(\mathsf{T}))$ is maximized by every $\sigma$ using transitions in $\mathbf{E}_n$ only hence also by every $\sigma$ using transitions in $\mathbf{E}_{n+1}$ only. The case with $v \in \mathsf{T}$ is trivial.

$\Leftarrow$: Let $\sigma$ be a greedily maximizing strategy on $n+1$ steps. If $\sigma$ used a transition $a \in \mathbf{E}_n \setminus \mathbf{E}_{n+1}$ in $v$ then it would not be greedily maximizing on $n+1$ steps, since the $n+1$ steps maximum (equal to $R_{n+1}(v)$) is not realized by $a$ as it has been cut off in the $(n+1)$-th step. Again similarly for the minimizing part. ☐

Since the number of transitions is finite, there is a fix-point $\mathbf{E}_n = \mathbf{E}_{n+1}$, moreover, $n \leq |\mathbf{E}|$. Therefore, any strategies using $\mathbf{E}_{|\mathbf{E}|}$ are greedily optimizing on $m$ steps for all $m$, hence greedily optimal. The complexity is thus polynomial in the size of the game graph.

As there is always a transition enabled in each vertex (the last one is trivially optimal), we can choose one transition in each vertex arbitrarily and thus get a greedy optimal stationary strategy. ☐

## C.8 Proof of Theorem 3.9

**THEOREM 3.9.** *Let $\sigma_g$ be a greedily maximizing stationary strategy, and $\pi_g$ a greedily minimizing stationary strategy. Let $\sigma$ be an optimal maximizing CD strategy, and $\pi$ an optimal minimizing CD strategy. Then for all sufficiently large $k \in \mathbb{N}$ we have that BCD strategies $\sigma' \in \Sigma$ and $\pi' \in \Pi$ defined by*

$$\sigma'(i,v) = \begin{cases} \sigma(i,v) & \textit{if } i < k; \\ \sigma_g(v) & \textit{otherwise.} \end{cases} \qquad \pi'(i,v) = \begin{cases} \pi(i,v) & \textit{if } i < k; \\ \pi_g(v) & \textit{otherwise.} \end{cases}$$

*are optimal. Moreover, if all transition probabilities in $\mathsf{G}$ are rational, then $\sigma'$ and $\pi'$ are optimal for all $k \geq rt(1 + m^{|A|^2 \cdot |V|^2})$, where $m$ is the maximal denominator of transition probabilities.*

PROOF. W.l.o.g. let us assume that all states in $\mathsf{T}$ are absorbing. Indeed, after reaching $\mathsf{T}$ any strategy can be applied without changing the $\mathcal{P}^{\sigma,\pi}_{v}(Reach^{\leq t}(\mathsf{T}))$. Also due to Theorem 3.3 and Theorem 3.6 we may restrict our attention to CD strategies. Therefore, in this proof, $\Sigma$ and $\Pi$ denote sets of CD strategies only.

Firstly, we prove a weaker version of the theorem for continuous-time Markov decision processes, and then extend it to games. From now on, let G be a finite game with one rate r and $V_\Diamond = \emptyset$. The case with $V_\Box = \emptyset$ is dual here. We write $\mathcal{P}^\sigma$ instead of $\mathcal{P}^{\sigma,\pi}$ where $\pi$ is the only strategy in $\Pi$.

Let us denote

$$\psi_n(t) \quad = \quad F_n(t) - F_{n+1}(t) = e^{-r\cdot t}\frac{(rt)^n}{n!}$$

the probability that by the time $t$ exactly $n$ subsequent actions take place. Using the notation from the proof of Lemma 3.8 in Section C.7 we have

$$\mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{\leq t}(T)) = \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$$

The following lemma proves a key argument that after enough time has elapsed, the probability of taking only one more step before the time limit is much larger than that of taking more than one.

**Lemma C.12.** *For every $\varepsilon > 0$ for all $n \geq rt(1+\frac{1}{\varepsilon})$*

$$\frac{F_{n+1}(t)}{\psi_n(t)} < \varepsilon$$

PROOF.

$$\frac{F_{n+1}(t)}{\psi_n(t)} = \frac{1}{n!}\sum_{i=1}^{\infty}\frac{(rt)^i}{(n+i)!} < \sum_{i=1}^{\infty}\frac{(rt)^i}{(n+1)^i} = \frac{rt}{n+1-rt} < \varepsilon$$

□

Nextly, we prove that any optimal strategy must eventually behave greedily, at least for a particular number of steps (recall the definition of greediness on $n$ steps from Section C.7). Subsequently, we prove that greediness on a large number of steps coincides with greediness. To cover the notion of finiteness of the game, we introduce *granularity* of the game to be the least common multiple of probabilities' denominators, i.e. we denote

$$M = \mathrm{lcm}\{r \mid \exists \alpha \in A, v \in V,\ q, r \in \mathbb{N} :\ \mathbf{P}(\alpha)(u) = \frac{q}{r},\ \gcd(q,r) = 1\}$$

**Lemma C.13.** *For every $n \in \mathbb{N}$ there is $\delta > 0$ such that for all $h \geq rt(1 + 1/\delta)$, for every optimal CD strategy $\sigma$, the strategy $\sigma[h]$ is greedily optimizing on $n$ steps. Moreover, if all transition probabilities are rational, then $\delta = (1/M)^n$.*

PROOF. Let $\sigma_g$ be greedily maximizing CD strategy and $\sigma$ an optimal CD strategy supposed, for a contradiction, not to be greedily maximizing. Then there is $j \leq n$ such that

$$\mathcal{P}^{\sigma_g}_{h,v}(\mathit{Reach}^{<\infty}_{\leq j}(T)) > \mathcal{P}^{\sigma}_{h,v}(\mathit{Reach}^{<\infty}_{\leq j}(T))$$

and for all $i < j$

$$\mathcal{P}^{\sigma_g}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) = \mathcal{P}^{\sigma}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$$

Therefore,

$$\begin{aligned}
&\mathcal{P}^{\sigma_g}_{h,v}(\mathit{Reach}^{\leq t}(T)) - \mathcal{P}^{\sigma}_{h,v}(\mathit{Reach}^{\leq t}(T)) = \\
&= \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}^{\sigma_g}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) - \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}^{\sigma}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) \geq \\
&\geq \psi_{h+j}(t)\mathcal{P}^{\sigma_g}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) - \psi_{h+j}(t)\mathcal{P}^{\sigma}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) - F_{h+j+1} = (*)
\end{aligned}$$

By the finiteness of the game there is $\delta > 0$ such that $\delta < \mathcal{P}^{\sigma_g}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) - \mathcal{P}^{\sigma}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$ for all $\sigma \in \Sigma$, $v \in V$ and $i \leq n$ whenever this difference is non-zero. Moreover, if all transition probabilities are rational, then $\delta$ can be chosen to be $(1/M)^n$. Indeed, $\mathcal{P}^{\tau}_{v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$ is clearly expressible as $\ell/M^i$ for some $\ell \in \mathbb{N}_0$. Hence,

$$(*) > \psi_{h+j}(t)\delta - F_{h+j+1}(t) > 0$$

by Lemma C.12 for all $h \geq rt(1 + 1/\delta)$, a contradiction with optimality of $\sigma$. □

**Lemma C.14.** *If a strategy is greedily optimizing on $|\mathbf{E}|$ steps, then it is greedily optimizing.*

PROOF. Using the notation of the algorithm in the proof of Lemma 3.8, $\mathbf{E}_{|\mathbf{E}|}$ contains exactly the transitions allowed for strategies that are greedily optimizing on $|\mathbf{E}|$ steps (by Lemma C.10). These are also exactly the transitions allowed for greedily optimizing strategies as $\mathbf{E}_{|\mathbf{E}|} = \mathbf{E}_{|\mathbf{E}|+i}$ for all $i \in \mathbb{N}$. □

**Corollary C.15.** *Let $k = rt(1 + M^{|\mathbf{E}|})$. Then for all $v \in V$ and $h \geq k$, for every optimal CD strategy $\sigma$, the strategy $\sigma[h]$ is greedily optimizing.* □

Note that since the number of probabilities is not greater than $|A| \cdot |V|$, we have $M \leq m^{|A|\cdot|V|}$, where $m$ is the greatest denominator of transition probabilities. Moreover, since $|\mathbf{E}| \leq |V| \cdot |A|$, we get $M^{|E|} \leq m^{|A|^2\cdot|V|^2}$. Furthermore, for probabilities encoded in binary, where $bp$ is the maximum bit-length, we get $m \leq 2^{bp}$ and thus $M \leq 2^{bp\cdot|A|\cdot|V|}$.

We now turn our attention to the proof for games in general. From now on, let $G$ be a uniform, i.e. with one rate, finite game. Let us recall $\sigma$ to be an optimal CD strategy,

$\sigma_g$ a stationary greedily maximizing strategy and $\sigma'$ such that $\sigma'(h, v) = \sigma_g(v)$ for all $h \geq k = rt(1 + M^{|\mathbf{E}|})$, and $\sigma' = \sigma$ otherwise, i.e. on the first $k$ steps. Then for all $v \in V$ and $h \geq k$

$$\min_{\pi \in \Pi} \mathcal{P}^{\sigma',\pi}_{h,v}(\mathit{Reach}^{\leq t}(T)) - \min_{\pi \in \Pi} \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{\leq t}(T)) =$$

$$= \min_{\pi \in \Pi} \sum_{i=0}^{\infty} \psi_{h+i}(t) \mathcal{P}^{\sigma_g,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) - \min_{\pi \in \Pi} \sum_{i=0}^{\infty} \psi_{h+i}(t) \mathcal{P}^{\sigma,\pi}_{h,v}(\mathit{Reach}^{<\infty}_{\leq i}(T)) = (*)$$

**Lemma C.16.** *For $\pi_g$ a greedily minimizing strategy*

$$\max_{\sigma \in \Sigma} \sum_{i=0}^{\infty} \psi_{k+i}(t) \mathcal{P}^{\sigma,\pi_g}_{v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$$

*is realized by all greedily maximizing strategies. Similarly, for $\sigma_g$ a greedily maximizing strategy*

$$\min_{\pi \in \Pi} \sum_{i=0}^{\infty} \psi_{k+i}(t) \mathcal{P}^{\sigma_g,\pi}_{v}(\mathit{Reach}^{<\infty}_{\leq i}(T))$$

*is realized by all greedily minimizing strategies.*

PROOF. Let $G(\hat{v}, -, \pi_g)$ denote the unfold of the game according to $\pi_g$ and leaving the $\Box$-nondeterminism thus resulting in a Markov decision process. By Corollary C.15 the maximum is realized by all greedily maximizing strategies in $G(\hat{v}, -, \pi_g)$. We prove that these are exactly the greedily minimizing strategies in G. Consider the algorithm from proof of Lemma 3.8 in Section C.7 applied on $G(\hat{v}, -, \pi_g)$ yielding $R'_i$ and $\mathbf{E}'_i$. We prove that $R_i = R'_i$ and $\mathbf{E}_i(v) = \mathbf{E}'_i(v)$ for all $i$ and $v \in V_\Box$.

Clearly, $R_0 = \mathbf{1}_T = R'_0$ and $\mathbf{E}_0(v) = \mathbf{E}(v) = \mathbf{E}'_0(v)$ for $v \in V_\Box$. On the one hand, for $v \in V_\Box$ the algorithm computes the same $R'_{i+1}(v) = R_{i+1}(v)$ and thus $\mathbf{E}'_{i+1}(v) = \mathbf{E}_{i+1}(v)$ using the induction hypothesis. On the other hand, $R'_{i+1}(v) = R_{i+1}(v)$ for $v \in V_\Diamond$ since the transition $\pi_g(v)$ realizes the minimum by greediness of $\pi_g$.

Therefore, we conclude again by characterization of greedy strategies in Lemma C.10 and the fix-point argument for $\mathbf{E}_{|\mathbf{E}|}$.

The minimizing part is dual. □

The first minimum in $(*)$ is realized by a greedily minimizing strategy $\pi_g \in \Pi$ by Lemma C.16 for minimizer. Hence

$$(*) = \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}_{h,v}^{\sigma_g,\pi_g}(Reach_{\leq i}^{<\infty}(T)) - \min_{\pi\in\Pi} \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}_{h,v}^{\sigma,\pi}(Reach_{\leq i}^{<\infty}(T)) \geq$$
$$\geq \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}_{h,v}^{\sigma_g,\pi_g}(Reach_{\leq i}^{<\infty}(T)) - \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}_{h,v}^{\sigma,\pi_g}(Reach_{\leq i}^{<\infty}(T)) \geq$$
$$\geq \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}_{h,v}^{\sigma_g,\pi_g}(Reach_{\leq i}^{<\infty}(T)) - \max_{\sigma\in\Sigma} \sum_{i=0}^{\infty} \psi_{h+i}(t)\mathcal{P}_{h,v}^{\sigma,\pi_g}(Reach_{\leq i}^{<\infty}(T)) = (**)$$

Considering $G(\hat{v}, \pi_g)$ and using Lemma C.16 again now for maximizer, $\sigma_g$ being greedily maximizing implies $(**) = 0$. Since $\sigma = \sigma'$ on the first $k$ steps, summing these inequalities for $h = k$ over all $v \in V$ (weighted by probabilities of being in $v$ after $h$ steps) yields

$$\min_{\pi\in\Pi} \mathcal{P}_{\hat{v}}^{\sigma',\pi}(Reach^{\leq t}(T)) - \min_{\pi\in\Pi} \mathcal{P}_{\hat{v}}^{\sigma,\pi}(Reach^{\leq t}(T)) \geq 0$$

Hence, $\sigma'$ is optimal. The minimizing part is similar. □

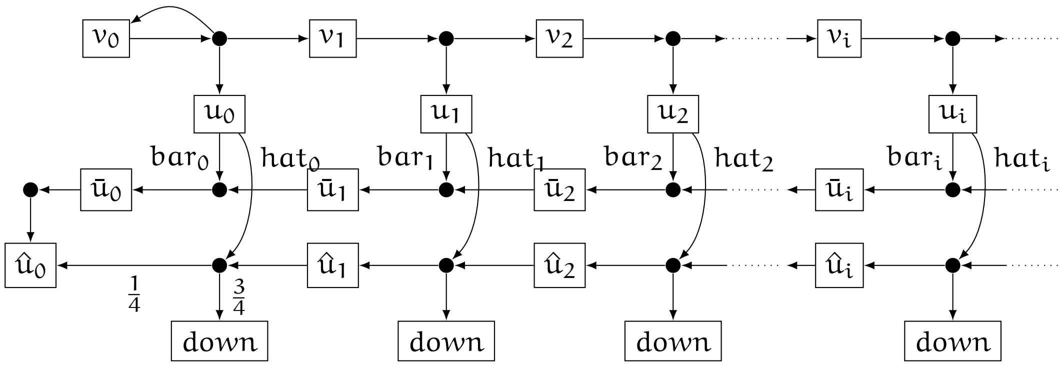
## C.9 Proof of Observation 3.10

**OBSERVATION 3.10.** *Optimal BCD strategies do not necessarily exist in infinite-state uniform CTGs, even if they are finitely-branching and use only finitely many distinct transition probabilities.*

PROOF. Consider a game $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$ where $V = V_\Box = \{v_i, u_i, \bar{u}_i, \hat{u}_i \mid i \in \mathbb{N}_0\} \cup \{down\}$, $A = \{a_i, hat_i, bar_i, \hat{b}_i, \bar{b}_i \mid \in \mathbb{N}_0\}$, $\mathbf{E}(v_i) = \{a_i\}$, $\mathbf{E}(u_i) = \{bar_i, hat_i\}$, $\mathbf{E}(\hat{u}_i) = \{\hat{b}_i\}$, and $\mathbf{E}(\bar{u}_i) = \{\bar{b}_i\}$ for all $i \in \mathbb{N}_0$. $\mathbf{P}$ is defined as follows:

- $\mathbf{P}(a_0)$ is the uniform distribution on $\{v_0, v_1, u_0\}$, $\mathbf{P}(a_i)$ where $i > 0$ is the uniform distribution on $\{u_i, v_{i+1}\}$,
- $\mathbf{P}(hat_i)(\hat{u}_i) = 1$ and $\mathbf{P}(bar_i)(\bar{u}_i) = 1$,
- $\mathbf{P}(\bar{b}_0)(\hat{u}_0) = 1$, and $\mathbf{P}(\bar{b}_j)(\bar{u}_{j-1}) = 1$ for $j > 0$,
- $\mathbf{P}(\hat{b}_1)(down) = \frac{3}{4}$, $\mathbf{P}(\hat{b}_1)(\hat{u}_0) = \frac{1}{4}$, and for $j > 1$ we set $\mathbf{P}(\hat{b}_j)$ to be the uniform distribution on $\{\hat{u}_{i-1}, down\}$.

We set $\mathbf{R}(a) = 1$ for all $a \in A$. The structure of $G$ is shown below.




Observe that player $\Box$ has a real choice only in states $u_i$.

We show that if the history is of the form $v_0 a_0 \ldots v_i a_i u_i$ (where $i \in \mathbb{N}_0$), the optimal strategy w.r.t. reaching $\hat{u}_0$ must choose the action $bar_i$. We need to show that $F_{2i+3}(1) > \frac{1}{2^{i+2}} \cdot F_{2i+2}(1)$, i.e., that $\frac{F_{2i+2}(1)}{F_{2i+3}(1)} < 2^{i+2}$. This follows by observing that

$$\frac{F_{2i+2}(1)}{F_{2i+3}(1)} = \frac{\sum_{j=2i+2}^{\infty} \frac{1}{j!}}{\sum_{j=2i+3}^{\infty} \frac{1}{j!}} < \frac{\sum_{j=2i+2}^{\infty} \frac{1}{j!}}{\frac{1}{(2i+3)!}} < (2i+3) + \sum_{k=0}^{\infty} \frac{1}{(2i+3)^k} = 2i+3+\frac{2i+4}{2i+3} < 2i+5 < 2^{i+2}\,.$$

On the other hand, from Corollary C.15 one can deduce that for all $i$ there is $j$ such that any optimal strategy must choose $hat_i$ if the history is of the form $(v_0 a_0)^j \ldots v_i a_i u_i$

$\Box$

# D Proofs of Section 4

## D.1 Proof of $\varepsilon$-optimality of the strategies $\sigma_\varepsilon$ and $\pi_\varepsilon$

**Lemma D.1.** *The strategies $\sigma_\varepsilon$ and $\pi_\varepsilon$ (defined in Section 4.1) are $\varepsilon$-optimal.*

PROOF. We use some notation of the proof of Theorem 3.1. Remember that given $\sigma \in \Sigma$, $\pi \in \Pi$, $\mathbf{j} \in \mathcal{H}$, and $u \in V$, we denote by $P^{\sigma,\pi}(u, \mathbf{j})$ the probability of all runs $\alpha \in Run_{G(u,\sigma,\pi)}$ such that for some $n \in \mathbb{N}_0$ the state $\alpha(n)$ hits $T$ and matches $\mathbf{j}$, and for all $0 \le j < n$ we have that $\alpha(j)$ does not hit $T$.

Given $(\sigma, \pi) \in \Sigma \times \Pi$, $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \le k$, and $v \in V$, we define

$$\bar{P}^{\sigma,\pi}(\mathbf{i}, v) := \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ |\mathbf{j}| \le k - |\mathbf{i}|}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j})$$

the probability of reaching $T$ from $v$ before time $t$ in at most $k - |\mathbf{i}|$ steps using the strategies $\sigma$ and $\pi$ and assuming that the history already matches $\mathbf{i}$. We have shown in the proof of Lemma C.7 that for every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$, the value

$$\max_{\sigma\in\Sigma}\min_{\pi\in\Pi} \bar{P}^{\sigma,\pi}(\mathbf{i}, v) = \min_{\pi\in\Pi}\max_{\sigma\in\Sigma} \bar{P}^{\sigma,\pi}(\mathbf{i}, v)$$

is equal to $\bar{R}(\mathbf{i}, v)$ defined by the following equations:

$$\bar{R}(\mathbf{i}, v) := \begin{cases} F_{\mathbf{i}}(t) & \text{if } v \in T \\ 0 & \text{if } v \notin T \text{ and } |\mathbf{i}| = k \\ \max_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_\Box \setminus T \text{ and } |\mathbf{i}| < k \\ \min_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_\Diamond \setminus T \text{ and } |\mathbf{i}| < k \end{cases}$$

Note that $\bar{P}^{\sigma,\pi}(\mathbf{0}, v) = \mathcal{P}^{\sigma,\pi}_v(\mathit{Reach}^{\leq t}_{\leq k}(T))$ and thus $\bar{R}(\mathbf{0}, v) = \mathit{val}^k(v)$, the k-step value in $v$.

Note that assuming $l_{\mathbf{i}}(t) = F_{\mathbf{i}}(t)$ for all $\mathbf{i} \in \mathcal{H}$ satisfying $|\mathbf{i}| \leq k$, we would obtain that each $R(\mathbf{i}, v)$ is precisely $\bar{R}(\mathbf{i}, v)$ and hence that $\sigma_\varepsilon$ and $\pi_\varepsilon$ are k-step optimal strategies.

Let us allow imprecisions in the computation of $l_{\mathbf{i}}(t)$. We proceed as follows: First we show, by induction, that each value $R(\mathbf{i}, v)$ approximates the value $\bar{R}(\mathbf{i}, v)$ with relative error $\frac{\varepsilon^{2|\mathbf{i}|+1}}{2^{2|\mathbf{i}|+1}}$ (Lemma D.2 below). From this we get, also by induction, that both $\min_{\pi\in\Pi} \bar{P}^{\sigma_\varepsilon,\pi}(\mathbf{i}, v)$ and $\max_{\sigma\in\Sigma} \bar{P}^{\sigma,\pi_\varepsilon}(\mathbf{i}, v)$ approximate $\bar{R}(\mathbf{i}, v)$ with relative error $\frac{\varepsilon^{2|\mathbf{i}|+1}}{2^{2|\mathbf{i}|+1}}$ as well (Lemma D.3 below). In other words, $\sigma_\varepsilon$ and $\pi_\varepsilon$ are $\frac{\varepsilon}{2}$-optimal strategies in the k-step game. Together with the assumptions imposed on k we obtain that $\sigma_\varepsilon$ and $\pi_\varepsilon$ are $\varepsilon$-optimal strategies.

We denote by $\mathit{err}_n$ the number $\frac{\varepsilon^{2n+1}}{2^{2n+1}}$.

**Lemma D.2.** *For all* $\mathbf{i} \in \mathcal{H}$ *and* $v \in V$ *we have*

$$(1 - \mathit{err}_{|\mathbf{i}|}) \cdot \bar{R}(\mathbf{i}, v) \quad \leq \quad R(\mathbf{i}, v) \quad \leq \quad (1 + \mathit{err}_{|\mathbf{i}|}) \cdot \bar{R}(\mathbf{i}, v)$$

PROOF. If $v \in T$, then $\bar{R}(\mathbf{i}, v) = F_{\mathbf{i}}(t)$ and $R(\mathbf{i}, v) = l_{\mathbf{i}}(t)$, and the inequality follows from the definition of $l_{\mathbf{i}}(t)$. Assume that $v \notin T$. We proceed by induction on $n = k - |\mathbf{i}|$. For $n = 0$ we have $\bar{R}(\mathbf{i}, v) = 0 = R(\mathbf{i}, v)$. Assume the inequality holds for any $v$ and $\mathbf{i} \in \mathcal{H}$ such that $|\mathbf{i}| = k - n$. Let us consider $\mathbf{i} \in \mathcal{H}$ such that $|\mathbf{i}| = k - n - 1$ and $v \in V$. If $v \in V_\Box$ we have

$$\begin{aligned} R(\mathbf{i}, v) &= \max_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\ &\leq \max_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot (1 + \mathit{err}_{|\mathbf{i}|+1}) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \cdot (1 + \mathit{err}_{|\mathbf{i}|+1}) \\ &= (1 + \mathit{err}_{|\mathbf{i}|+1})^2 \cdot \max_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) \\ &\leq (1 + \mathit{err}_{|\mathbf{i}|}) \cdot \bar{R}(\mathbf{i}, v) \end{aligned}$$

and, similarly,

$$R(\mathbf{i}, v) \geq (1 - err_{|\mathbf{i}|}) \cdot \bar{R}(\mathbf{i}, v)$$

For $v \in V_\Diamond$ the proof is similar. $\square$

We denote by $\Sigma_{CD}$ and $\Pi_{CD}$ the sets of all CD strategies of $\Sigma$ and $\Pi$, respectively. Given a strategy $\tau \in \Sigma_{CD} \cup \Pi_{CD}$ and $\mathbf{i} \in \mathcal{H}$, we denote by $\tau[\mathbf{i}]$ the strategy obtained from $\tau$ by $\tau[\mathbf{i}](\mathbf{j}, u) := \tau(\mathbf{i} + \mathbf{j}, u)$.

Given $\mathbf{i} \in \mathcal{H}$ and $\pi \in \Pi$, we define

$$K^\pi(\mathbf{i}, v) := \bar{P}^{\sigma_\varepsilon[\mathbf{i}], \pi[\mathbf{i}]}(\mathbf{i}, v)$$

Similarly, given $\mathbf{i} \in \mathcal{H}$ and $\sigma \in \Sigma$, we define

$$K^\sigma(\mathbf{i}, v) := \bar{P}^{\sigma[\mathbf{i}], \pi_\varepsilon[\mathbf{i}]}(\mathbf{i}, v)$$

**Lemma D.3.** *Let $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$. We have*

$$\min_{\pi \in \Pi_{CD}} K^\pi(\mathbf{i}, v) \geq \bar{R}(\mathbf{i}, v) \cdot (1 - err_{|\mathbf{i}|})$$
$$\max_{\sigma \in \Sigma_{CD}} K^\sigma(\mathbf{i}, v) \leq \bar{R}(\mathbf{i}, v) \cdot (1 + err_{|\mathbf{i}|})$$

PROOF. If $v \in T$, then $K^\pi(\mathbf{i}, v) = K^\sigma(\mathbf{i}, v) = F_\mathbf{i}(t)$ and $\bar{R}(\mathbf{i}, v) = l_\mathbf{i}(t)$, and similarly as above, the result follows from the definition of $l_\mathbf{i}(t)$. Assume that $v \notin T$. We proceed by induction on $n := k - |\mathbf{i}|$. For $n = 0$ we have $0 = K^\pi(\mathbf{i}, v) = K^\sigma(\mathbf{i}, v) = \bar{R}(\mathbf{i}, u)$. Assume the lemma holds true for $n$ and consider $n + 1$. If $v \in V_\Box$ and $\sigma_\varepsilon(\mathbf{i}, v)(b) = 1$,

$$\begin{aligned}
\min_{\pi \in \Pi_{CD}} K^\pi(\mathbf{i}, v) &= \min_{\pi \in \Pi_{CD}} \sum_{u \in V} \mathbf{P}(b)(u) \cdot K^\pi(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \\
&= \sum_{u \in V} \mathbf{P}(b)(u) \cdot \min_{\pi \in \Pi_{CD}} K^\pi(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \\
&\geq \sum_{u \in V} \mathbf{P}(b)(u) \cdot \bar{R}(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \cdot (1 - err_{|\mathbf{i}|+1}) \\
&\geq \sum_{u \in V} \mathbf{p}(b)(u) \cdot \frac{1}{1 + err_{|\mathbf{i}|+1}} \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(b)}, u) \cdot \frac{1 - err_{|\mathbf{i}|+1}}{1 + err_{|\mathbf{i}|+1}} \\
&= R(\mathbf{i}, v) \cdot \frac{1 - err_{|\mathbf{i}|+1}}{(1 + err_{|\mathbf{i}|+1})^2} \\
&\geq \bar{R}(\mathbf{i}, v) \cdot (1 - err_{|\mathbf{i}|+1}) \cdot \frac{1 - err_{|\mathbf{i}|+1}}{(1 + err_{|\mathbf{i}|+1})^2} \\
&\geq \bar{R}(\mathbf{i}, v) \cdot (1 - err_{|\mathbf{i}|})
\end{aligned}$$

and

$$
\begin{aligned}
\max_{\sigma\in\Sigma_{CD}} K^{\sigma}(\mathbf{i},v) &= \max_{\sigma\in\Sigma} \sum_{a\in\mathbf{E}(v)} \sigma(\mathbf{i},v)(a) \sum_{u\in V} \mathbf{P}(a)(u)\cdot K^{\sigma}(\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)},u) \\
&= \max_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u)\cdot \max_{\sigma\in\Sigma} K^{\sigma}(\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)},u) \\
&\le \max_{a\in\mathbf{E}(u)} \sum_{u\in V} \mathbf{P}(a)(u)\cdot \bar{R}(\mathbf{i}+\mathbf{1}_{\mathbf{R}(a)},u)\cdot(1+\mathit{err}_{|\mathbf{i}|+1}) \\
&\le \bar{R}(\mathbf{i},u)\cdot(1+\mathit{err}_{|\mathbf{i}|})\,.
\end{aligned}
$$

For $u \in V_{\Diamond}$ the proof is similar. □

## D.2 Proof of Theorem 4.1

**THEOREM 4.1.** *Assume that* $G$ *is finite. Then for every* $\varepsilon > 0$ *there are* $\varepsilon$*-optimal BCD strategies* $\sigma_\varepsilon \in \Sigma$ *and* $\pi_\varepsilon \in \Pi$ *computable in time* $|V|^2 \cdot |A| \cdot bp \cdot (|\mathcal{R}|+1)^{\mathcal{O}((\max\mathcal{R})\cdot t+\ln\frac{1}{\varepsilon})}$.

First, we show that the phase 1. takes time exponential w.r.t. $k$ and polynomial w.r.t. $\frac{1}{\varepsilon}$. We approximate the value of $F_{\mathbf{i}}(t)$ to the relative precision $(\varepsilon/2)^{2k+1}$ as follows. According to [1], the value of $F_{\mathbf{i}}(t)$ is expressible as $\sum_{r\in\mathcal{R}} q_r e^{-rt}$, where $q_r$ is a polynomial in $t$ and can be precisely computed using exponentially many (in $k$) arithmetical operations on polynomially large integers, hence is in **EXPTIME** w.r.t. $k$, i.e. $(\max\mathcal{R})t + \ln\frac{1}{\varepsilon}$. We approximate this fraction with a floating point representation with relative error $(\varepsilon/4)^{2k+1}$. This can be done in linear time w.r.t. the length of numbers and $k\ln\frac{1}{\varepsilon}$, hence in the same exponential time w.r.t. $k$.

For every $r \in \mathcal{R}$, the value of $e^{-rt}$ can be approximated using Taylor's theorem. After the $n$-th summand is computed, the remainder is smaller than $\left(\frac{ert}{n}\right)^n$ using Stirling's formula. For $n = k(4/\varepsilon) > 4(\max\mathcal{R})te^2/\varepsilon$ we get the relative error to be less than $\left(\frac{\varepsilon}{4}\right)^n < \left(\frac{\varepsilon}{4}\right)^{2k+1}$. To compute the $n$-th approximation, time polynomial in $n$ is sufficient. Hence, its complexity is polynomial w.r.t. $k/\varepsilon$, i.e. $((\max\mathcal{R})t + \ln\frac{1}{\varepsilon})/\varepsilon$ and linear w.r.t. $bp$.

Note that the above procedure has to be repeated for every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \le k$. The number of all $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \le k$, is $\binom{|\mathcal{R}|+k-1}{k} \le |\mathcal{R}|^k$. So computing all values $F_{\mathbf{i}}(t)$ takes time $|\mathcal{R}|^k \cdot |\mathcal{R}| \cdot bp \cdot \left(\frac{1}{\varepsilon}\right)^{\mathcal{O}(1)} \cdot 2^{\mathcal{O}((\max\mathcal{R})\cdot t+\ln\frac{1}{\varepsilon})}$.

Using similar procedure as above, for every $a \in A$ and $u \in V$, we compute the floating point approximation $\mathbf{p}(a)(u)$ of $\mathbf{P}(a)(u)$ to the relative precision $(\varepsilon/4)^{2k+1}$ in time linear in $k\ln\frac{1}{\varepsilon}$ and $bp$. (Here we assume that the probabilities $\mathbf{P}(a)(u)$ are given as fractions with both numerator and denominator represented in binary with length bounded by *bp*.)

So the first phase takes time $|\mathcal{R}|^k \cdot |A| \cdot |V| \cdot |\mathcal{R}| \cdot \left(\frac{1}{\varepsilon}\right)^{\mathcal{O}(1)} \cdot bp \cdot 2^{\mathcal{O}((\max \mathcal{R}) \cdot t + \ln \frac{1}{\varepsilon})}$.

Second, we evaluate the complexity of phase 2. In phase 2., the algorithm computes the table R and outputs the results into the table C. The complexity is thus determined by the product of the table size and the time to compute one item in the table. The size of the tables is $\binom{|\mathcal{R}|+k-1}{k} \cdot |V| \leq |\mathcal{R}|^k |V|$.

The value of $R(\mathbf{i}, u)$ according to the first case has already been computed in phase 1. To compute the value according to the third or fourth case we have to compare numbers whose representation has at most $\left(\frac{1}{\varepsilon}\right)^{\mathcal{O}(1)} 2^{\mathcal{O}(k)} + k \cdot k \ln(\frac{1}{\varepsilon}) \cdot bp \cdot |V|$ bits. To compute $R(\mathbf{i}, v)$, we need to compare $|A|$ such numbers. So the phase 2. takes time $|\mathcal{R}|^k |V| \cdot |A| \cdot (\left(\frac{1}{\varepsilon}\right)^{\mathcal{O}(1)} 2^{\mathcal{O}(k)} + k \cdot k \ln(\frac{1}{\varepsilon}) \cdot bp \cdot |V|)$.

Altogether, the overall complexity is

$$|V|^2 \cdot |A| \cdot |\mathcal{R}|^{\mathcal{O}(k)} \cdot \left(\frac{1}{\varepsilon}\right)^{\mathcal{O}(1)} \cdot bp \cdot 2^{\mathcal{O}(k)} = |V|^2 \cdot |A| \cdot bp \cdot (|\mathcal{R}| + 1)^{\mathcal{O}((\max \mathcal{R})t + \ln \frac{1}{\varepsilon})}$$

□

## D.3 Proof of Corollary 4.2

**COROLLARY 4.2.** *Let* G *be a finitely-branching game with bounded rates and let* $v \in V$. *Assume that the vertices and actions of* G *reachable from* $v$ *in a given finite number of steps are effectively computable, and that an upper bound on rates is also effectively computable. Then for every* $\varepsilon > 0$ *there are effectively computable BCD strategies* $\sigma_\varepsilon \in \Sigma$ *and* $\pi_\varepsilon \in \Pi$ *that are* $\varepsilon$*-optimal in* $v$.

PROOF. By Lemma 3.5, there is $k \in \mathbb{N}$ such that all k-step optimal strategies are $\frac{\varepsilon}{4}$-optimal. Thus we may safely restrict the set of vertices of the game G to the set $V_{reach}$ of vertices reachable from $v$ in at most k steps (i.e. for all $v' \in V_{reach}$ there is a sequence $v_0 \ldots v_k \in V^*$ and $a_0 \ldots a_k \in A^*$ such that, $v_0 = v$, $v_n = v'$, $a_i \in \mathbf{E}(v_i)$ for all $0 \leq i \leq n$ and $\mathbf{P}(a_i)(v_{i+1}) > 0$ for all $0 \leq i < n$). Moreover, for every action $a \in A$ which is enabled in a vertex of $V_{reach}$ there is a finite set $B_a$ of vertices such that $1 - \sum_{u \in B_a} \mathbf{P}(a)(u) < \frac{\varepsilon}{4k}$. We restrict the domain of $\mathbf{P}(a)$ to $B_a$ by assigning the probability 0 to all vertices of $V \setminus B_a$ and adding the probability $1 - \sum_{u \in B_a} \mathbf{P}(a)(u)$ to an arbitrary vertex of $B_a$. Finally, we restrict the set of vertices once more to the vertices reachable in k steps from $v$ using the restricted $\mathbf{P}$. Then the resulting game is finite and by Theorem 4.1 there is an $\frac{\varepsilon}{4}$-optimal

BCD strategy $\sigma'$ in this game. Now it suffices to extend $\sigma'$ to a BCD strategy $\sigma$ in the original game by defining, arbitrarily, its values for vertices and actions removed by the above procedure. It is easy to see that $\sigma$ is an $\varepsilon$-optimal BCD strategy in G. □

## D.4 Proof of Theorem 4.3

**THEOREM 4.3.** *The BCD strategies $\sigma_{max}$ and $\pi_{min}$ are optimal and effectively computable.*

PROOF. We start by showing that $\sigma_{max}$ and $\pi_{min}$ are optimal. Let us denote by $\Sigma_g$ (resp. $\Pi_g$) the set of all CD strategies $\sigma \in \Sigma$ (resp. $\pi \in \Pi$) such that for all $u \in V_\square$ ($u \in V_\Diamond$) and $i \geq k$ we have $\sigma(i, u) = \sigma_g(u)$. By Theorem 3.9, for every $v \in V$ we have

$$val(v) = \max_{\sigma \in \Sigma_g} \min_{\pi \in \Pi_g} \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}(T)) = \min_{\pi \in \Pi_g} \max_{\sigma \in \Sigma_g} \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}(T))$$

Let us denote

$$\bar{P}^{\sigma,\pi}(i, v) = \sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma,\pi}(Reach_{=j}^{<\infty}(T))$$

(This corresponds to $\bar{P}^{\sigma,\pi}(\mathbf{i}, v)$, as defined in the proof of Lemma 3.5 (2.), for uniform games).

For every $i \geq 0$ we put

$$val(i, v) = \max_{\sigma \in \Sigma_g} \min_{\pi \in \Pi_g} \bar{P}^{\sigma,\pi}(i, v) = \min_{\pi \in \Pi_g} \max_{\sigma \in \Sigma_g} \bar{P}^{\sigma,\pi}(i, v) \tag{5}$$

(Here the second equality follows from Theorem 3.1.)

Remember that given a CD strategy $\tau$ and $i \geq 0$, we denote by $\tau[i]$ a strategy obtained from $\tau$ by $\tau[i](j, u) = \tau(i + j, u)$. We denote by $\Sigma_{CD}$ and $\Pi_{CD}$ the sets of all CD strategies of $\Sigma$ and $\Pi$, respectively.

Given $i \geq 0$ and $\pi \in \Pi$, we define

$$\bar{K}^{\pi}(i, v) := \bar{P}^{\sigma_{max}[i],\pi[i]}(i, v)$$

Similarly, given $i \in \mathcal{H}$ and $\sigma \in \Sigma$, we define

$$\bar{K}^{\sigma}(i, v) := \bar{P}^{\sigma[i],\pi_{min}[i]}(i, v)$$

**Lemma D.7.** *Let $i \leq k$ and $v \in V$. We have*

$$\min_{\pi \in \Pi_{CD}} \bar{K}^{\pi}(i, v) = R(i, v) = \max_{\sigma \in \Sigma_{CD}} \bar{K}^{\sigma}(i, v) \tag{6}$$

$$R(i, v) = val(i, v) \tag{7}$$

PROOF. We start by proving the equation (6). If $v \in T$, then $K^{\pi}(i, v) = K^{\sigma}(i, v) = F_i(t) = R(i, v)$. Assume that $v \notin T$. We proceed by induction on $n = k - i$. For $n = 0$ we have

$$\bar{K}^{\pi}(i, v) = \bar{K}^{\sigma}(i, v) = \bar{P}^{\sigma_g, \pi_g}(i, v) = R(i, v)$$

Assume the lemma holds true for $n$ and consider $n + 1$. If $v \in V_{\Box}$ and $\sigma_{max}(\mathbf{i}, v)(b) = 1$,

$$\begin{aligned}
\min_{\pi \in \Pi_{CD}} \bar{K}^{\pi}(i, v) &= \min_{\pi \in \Pi_{CD}} \sum_{u \in V} \mathbf{P}(b)(u) \cdot \bar{K}^{\pi}(i + 1, u) \\
&= \sum_{u \in V} \mathbf{P}(b)(u) \cdot \min_{\pi \in \Pi_{CD}} \bar{K}^{\pi}(i + 1, u) \\
&= \sum_{u \in V} \mathbf{P}(b)(u) \cdot R(i + 1, u) \\
&= \max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i + 1, u) \\
&= R(i, v)
\end{aligned}$$

and

$$\begin{aligned}
\max_{\sigma \in \Sigma_{CD}} \bar{K}^{\sigma}(i, v) &= \max_{\sigma \in \Sigma} \sum_{a \in \mathbf{E}(v)} \sigma(i, v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \cdot \bar{K}^{\sigma}(i + 1, u) \\
&= \max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot \max_{\sigma \in \Sigma} \bar{K}^{\sigma}(i + 1, u) \\
&= \max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i + 1, u) \\
&= R(i, u)
\end{aligned}$$

For $u \in V_{\Diamond}$ the proof is similar.

Now the equation (7) follows easily:

$$R(i, v) = \min_{\pi \in \Pi_{CD}} \bar{K}^{\pi}(i, v) \leq \max_{\sigma \in \Sigma_{CD}} \min_{\pi \in \Pi_{CD}} \bar{P}^{\sigma, \pi}(i, v) =$$
$$\min_{\pi \in \Pi_{CD}} \max_{\sigma \in \Sigma_{CD}} \bar{P}^{\sigma, \pi}(i, v) \leq \max_{\sigma \in \Sigma_{CD}} \bar{K}^{\sigma}(i, v) = R(i, v)$$

□

This proves that $\sigma_{max}$ and $\pi_{min}$ are optimal.

**Effective computability of $\sigma_{max}$ and $\pi_{min}$.** We show how to compute the table $C(i,v)$. Assume that we have already computed the symbolic representations of the values $R(i+1,u)$ for all $u \in V$. Later we show that $\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_{v}^{\sigma_g,\pi_g}(\mathit{Reach}_{=j}^{<\infty}(T))$ can effectively be expressed as a linear combination of transcendental numbers of the form $e^{ct}$ where $c$ is algebraic. Therefore, for $u, u' \in V$ we have that $R(i+1,u) - R(i+1,u')$ can effectively be expressed as a finite sum $\sum_j \eta_j e^{\delta_j}$ where the $\eta_j$ and $\delta_j$ are algebraic complex numbers and the $\delta_j$'s are pairwise distinct. Now it suffices to apply Lemma 2. of [2] to decide whether $R(i+1,u) - R(i+1,u') > 0$, or not.

It remains to show that $\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_{v}^{\sigma_g,\pi_g}(\mathit{Reach}_{=j}^{<\infty}(T))$ is effectively expressible in the form $\sum_j \eta_j e^{\delta_j}$. Consider a game $G'$ obtained from $G$ by adding new vertices $v_1, \ldots, v_i$ and new actions $a_1, \ldots, a_i$, setting $\mathbf{E}(v_j) = \{a_j\}$ for $0 \le j \le i$, and setting $\mathbf{P}(a_i)(v) = 1$, and $\mathbf{P}(a_j)(v_{j+1}) = 1$ for $0 \le j < i$ (intuitively, we have just added a simple path of length $i$ from a new vertex $v_1$ to $v$). We put $\mathbf{R}(a_j) = r$ for $0 \le j \le i$. As the strategies $\sigma_g$ and $\pi_g$ are stationary, they can be used in $G'$ (we just make them select $a_j$ in $v_j$).

Since $v_j \notin T$ for all $0 \le j \le i$ we obtain

$$\mathcal{P}_{v_1}^{\sigma_g,\pi_g}(\mathit{Reach}^{\le t}(T)) = \sum_{j=0}^{\infty} F_j(t) \cdot \mathcal{P}_{v_1}^{\sigma_g,\pi_g}(\mathit{Reach}_{=j}^{<\infty}(T)) =$$
$$\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_{v_1}^{\sigma_g,\pi_g}(\mathit{Reach}_{=i+j}^{<\infty}(T)) = \sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_{v}^{\sigma_g,\pi_g}(\mathit{Reach}_{=j}^{<\infty}(T))$$

As $\sigma_g$ and $\pi_g$ are stationary, the chain $G'(v_1, \sigma_g, \pi_g)$ can be treated as a finite continuous time Markov chain. Therefore we may apply results of [1] and obtain the desired form of $\mathcal{P}_{v_1}^{\sigma_g,\pi_g}(\mathit{Reach}^{\le t}(T))$, and hence also of $\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_{v}^{\sigma_g,\pi_g}(\mathit{Reach}_{=j}^{<\infty}(T))$.