Continuous-Time Stochastic Games with Time-Bounded Reachability[☆]

Tomáš Brázdil, Vojtěch Forejt[1], Jan Krčál, Jan Křetínský[2,*], Antonín Kučera

Faculty of Informatics, Masaryk University, Botanická 68a, 60200 Brno, Czech Republic

Abstract

We study continuous-time stochastic games with time-bounded reachability objectives and time-abstract strategies. 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. Further, we show how to compute ε-optimal strategies in finite games and provide detailed complexity estimations. Moreover, we show how to compute ε-optimal strategies in infinite games with finite branching and bounded rates where the bound as well as the successors of a given state are effectively computable. Finally, we show how to compute optimal strategies in finite uniform games.

Keywords: continuous time stochastic systems, time-bounded reachability, stochastic games

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

[☆]This is an extended version of FSTTCS'09 paper with full proofs and improved complexity bounds on the ε-optimal strategies algorithm. The work has been supported by Czech Science Foundation, grant No. P202/10/1469. Jan Křetínský is a holder of Brno PhD Talent Financial Aid. Vojtěch Forejt was supported in part by ERC Advanced Grant VERIWARE and a Royal Society Newton Fellowship.

[*]Corresponding author, phone no.: +49 89 289 17236 , fax no.: +49 89 289 17207

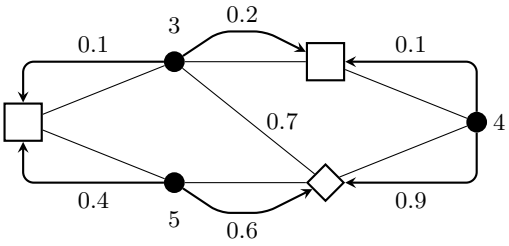
Email addresses: brazdil@fi.muni.cz (Tomáš Brázdil), vojfor@cs.ox.ac.uk (Vojtěch Forejt), krcal@fi.muni.cz (Jan Krčál), jan.kretinsky@fi.muni.cz (Jan Křetínský), kucera@fi.muni.cz (Antonín Kučera)

[1]Present address: Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK

[2]Present address: Institut für Informatik, Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany

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., [**?** **?** **?**] for more comprehensive expositions of results related to games in formal analysis and verification of 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., [**?** **?**]).

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 in each vertex 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 [**?** **?**]. 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 [**?**], 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 [**?**] is whether this result can be extended to all (not necessarily uniform) CTMDP. In [**?**], it is also shown that time-dependent strategies are more powerful than time-abstract ones, and this issue is addressed in greater detail in [**?**] where the mutual relationship between various classes of time-dependent strategies in CTMDPs is studied. Furthermore, in [**?**] 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 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 $\sup_\sigma \inf_\pi \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$ and $\inf_\pi \sup_\sigma \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$, where σ and π range over all time-abstract strategies of player $\Box$ and player $\Diamond$, and $\mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$ is the probability of reaching T before time t when starting in v in a play obtained by applying the strategies σ and π. 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"). Using the technique of uniformization, one can generalize this result to all finite (not necessarily uniform) games, see [**?**].

Algorithms. We show that for finite CTGs, ε-optimal strategies for both players are computable in $|V|^2 \cdot |A| \cdot bp^2 \cdot \left((\max \mathcal{R}) \cdot t + \ln \frac{1}{\varepsilon}\right)^{2|\mathcal{R}|+\mathcal{O}(1)}$ 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 [**?**] (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 5.26). 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.

Some proofs that are rather technical have been shifted into Appendix C.

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 *Dirac* if $f(a) = 1$ for some $a \in A$. The set of all distributions on A is denoted by $\mathcal{D}(A)$. A *σ-field* over a set Ω is a set $\mathcal{F} \subseteq 2^{\Omega}$ that contains Ω and is closed under complement and countable union. A *measurable space* is a pair $(\Omega, \mathcal{F})$ where Ω is a set called *sample space* and $\mathcal{F}$ is a σ-field over Ω 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., [**?**]). 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 $\mathcal{M} = (S, \mathbf{P}, \mathbf{R}, \mu)$, where S is a finite or countably infinite set of *states*, $\mathbf{P}$ is a *transition probability function* assigning to each $s \in S$ a probability distribution over S, $\mathbf{R}$ is a function assigning to each $s \in S$ a positive real *rate*, and μ is the *initial probability distribution* on S.

If $\mathbf{P}(s)(s') = x > 0$, we write $s \xrightarrow{x} s'$ or shortly $s \to s'$. A *time-abstract path* is a finite or infinite sequence $u = u_0, u_1, \ldots$ of states such that $u_{i-1} \to u_i$ for every $1 \leq i < \mathit{length}(u)$, where $\mathit{length}(u)$ is the length of u (the length of

an infinite sequence is ∞). 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 $\mathit{length}(u) = \mathit{length}(I)+1$. 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 σ-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

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

Note that if $\mathit{length}(u) = 1$, 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 [**?** **?**] 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 even if G is finitely branching). 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 N is 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 [**?** **?**], 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 real 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 $wv \in N^*V_\odot$ assigns a probability distribution on $\mathbf{E}(v)$. The sets of all strategies for player $\Box$ and player $\Diamond$ are denoted by Σ and Π, 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 N^*A is the set of states, the rate of a given $wa \in N^*A$ is $\mathbf{R}(a)$ (the rate function of $G(\hat{v}, \sigma, \pi)$ is also denoted by $\mathbf{R}$), and the only non-zero transition probabilities are between states of the form wa and $wava'$ with $wa \xrightarrow{x} wava'$ iff one of the following conditions is satisfied:

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

The initial distribution is determined as follows:

- $\mu(\hat{v}a) = \sigma(\hat{v})(a)$ if $\hat{v} \in V_\Box$ and $a \in \mathbf{E}(\hat{v})$;
- $\mu(\hat{v}a) = \pi(\hat{v})(a)$ if $\hat{v} \in V_\Diamond$ and $a \in \mathbf{E}(\hat{v})$;
- in the other cases, μ 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 $\mu(\hat{v}a) > 0$, 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, $v_1a_1v_2a_2$ 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 $wa \in N^*A$.

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 $wx \in N^*N$ we

define the vector $\mathbf{i}_{wx} \in \mathcal{H}$ such that $\mathbf{i}_{wx}(r)$ returns the cardinality of the set $\{j \in \mathbb{N}_0 \mid 0 \leq j < \mathit{length}(w), w(j) \in A, \mathbf{R}(w(j)) = r\}$. (Note that the last element x of wx is disregarded.) Given $\mathbf{i} \in \mathcal{H}$ and $wx \in N^*N$, we say that wx *matches* $\mathbf{i}$ if $\mathbf{i} = \mathbf{i}_{wx}$.

We say that a strategy τ is *counting (C)* if $\tau(wv) = \tau(w'v)$ for all $v \in V$ and $w, w' \in N^*$ such that $\mathbf{i}_{wv} = \mathbf{i}_{w'v}$. In other words, a strategy τ is counting if the only information about the history of a play w which influences the decision of τ is the vector $\mathbf{i}_{wv}$. Hence, every counting strategy τ 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 τ is *bounded counting (BC)* if there is $k \in \mathbb{N}$ such that $\tau(wv) = \tau(w'v)$ whenever $\mathit{length}(w) \geq k$ and $\mathit{length}(w') \geq k$. A strategy τ 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* $t \in \mathbb{R}^{>0}$. Let v be an initial vertex. Then each pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ determines a unique *outcome* $\mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$, which is the probability of all $\alpha \in \mathit{Run}_{G(v,\sigma,\pi)}$ that visit T before time t (i.e., there is $i \in \mathbb{N}_0$ such that $\alpha(i)$ hits T and $\sum_{j=0}^{i-1} \alpha[j] \leq t$). The goal of player $\Box$ is to maximize the outcome, while player $\Diamond$ aims at the opposite. In our next definition we recall the standard concept of an equilibrium outcome called the *value*.

Definition 2.3. We say that a vertex $v \in V$ *has a value* if

$$\sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T)) \quad = \quad \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T))$$

If v has a value, then $\mathit{val}(v)$ denotes the *value of* v defined by the above equality.

The existence of $\mathit{val}(v)$ follows easily by applying the powerful result of Martin about weak determinacy of Blackwell games [**?**] (more precisely, one can use the determinacy result for stochastic games presented in [**?**] which builds on [**?**]). In Section 3, we give a self-contained proof of the existence of $\mathit{val}(v)$, which also brings further insights used later in our algorithms. Still, we think it is worth noting how the existence of $\mathit{val}(v)$ follows from the results of [**?** **?**] because the argument is generic and can be used also for more complicated timed objectives and a more general class of games over *semi-Markov processes* [**?**] where the distribution of time spent by performing a given action is not necessarily exponential.

Theorem 2.4. *Every vertex $v \in V$ has a value.*

Proof. Let us consider an infinite path of G initiated in v, i.e., an infinite sequence $v_0, a_0, v_1, a_1, \ldots$ where $v_0 = v$ and $a_i \in \mathbf{E}(v_i)$, $\mathbf{P}(a_i)(v_{i+1}) > 0$ for all

$i \in \mathbb{N}_0$. Let f be a real-valued function over infinite paths of G defined as follows:

- If a given path does not visit a target vertex (i.e., $v_i \notin T$ for all $i \in \mathbb{N}_0$), then f returns 0;
- otherwise, let $i \in \mathbb{N}_0$ be the least index such that $v_i \in T$. The function f returns the probability $\mathcal{P}(X_0 + \cdots + X_{i-1} \leq t)$ where every X_j, $0 \leq j < i$, is an exponentially distributed random variable with the rate $\mathbf{R}(a_j)$ (we assume that all X_j are mutually independent). Intuitively, f returns the probability that the considered path reaches v_i before time t.

Note that f is Borel measurable and bounded. Also note that every run in a play $G(v, \sigma, \pi)$ initiated in v determines exactly one infinite path in G (the time stamps are ignored). Hence, f determines a unique random variable over the runs in $G(v, \sigma, \pi)$ which is denoted by $f_v^{\sigma,\pi}$. Observe that $f_v^{\sigma,\pi}$ does not depend on the time stamps which appear in the runs of $G(v, \sigma, \pi)$, and hence we can apply the results of [**?**] and conclude that

$$\sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathbb{E}[f_v^{\sigma,\pi}] \quad = \quad \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \mathbb{E}[f_v^{\sigma,\pi}]$$

where $\mathbb{E}[f_v^{\sigma,\pi}]$ is the expected value of $f_v^{\sigma,\pi}$. To conclude the proof, it suffices to realize that $\mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}^{\leq t}(T)) = \mathbb{E}[f_v^{\sigma,\pi}]$. □

Since values exist, it makes sense to define *ε-optimal* and *optimal* strategies.

Definition 2.5. Let $\varepsilon \geq 0$. We say that a strategy $\sigma \in \Sigma$ is an *ε-optimal maximizing strategy in v* (or just *ε-optimal in v*) if

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

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

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

A strategy is *ε-optimal* if it is ε-optimal in every v. A strategy is *optimal in v* if it is 0-optimal in v, and just *optimal* if it is optimal in every v.

3. The Existence of Values and Optimal Strategies

In this section we first give a self-contained proof that every vertex in a CTG with time-bounded reachability objectives has a value (Theorem 3.6). The argument does not require any additional restrictions, i.e., it works also for CTGs with infinite state-space and infinite branching degree. As we shall see, the ideas presented in the proof of Theorem 3.6 are useful also for designing an algorithm which for a given $\varepsilon > 0$ computes ε-optimal strategies for both

players. 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 $X_{\mathbf{i}} = \sum_{r \in \mathcal{R}} \sum_{i=1}^{\mathbf{i}(r)} X_i^{(r)}$ 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_ℓ to denote $F_{\mathbf{i}}$ 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 $\mathcal{P}_v^{\sigma,\pi}(\mathit{Reach}_{\sim k}^{\leq t}(T))$ the probability of all $\alpha \in \mathit{Run}_{G(v,\sigma,\pi)}$ that visit T for the first time in the number of steps satisfying the constraint $\sim k$ 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 $\sum_{j=0}^{i-1} \alpha[j] \leq t$).

We first restate Theorem 2.4 and give its constructive proof.

Theorem 3.6. *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 $\alpha \in \mathit{Run}_{G(u,\sigma,\pi)}$ 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

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

Clearly, it suffices to prove that $\mathcal{A} = \mathcal{B}$, because then for every vertex $v \in V$ we also have that $\mathcal{A}(\mathbf{0}, v) = \mathcal{B}(\mathbf{0}, v) = \mathit{val}(v)$. 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 $\mathcal{A} \preceq \mathcal{B}$ (where $\preceq$ is the standard pointwise order) is obvious and follows directly from the definition of $\mathcal{A}$ and $\mathcal{B}$. Hence, it suffices to prove the following two assertions:

1. By the following claim we obtain $\mu\mathcal{V} \preceq \mathcal{A}$.

   Claim 3.7. $\mathcal{A}$ is a fixed point of $\mathcal{V}$.

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 which we get $\mathcal{B} \preceq \mu\mathcal{V}$.

   The strategy π_ε 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 π_ε indeed satisfies the above equality. 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 π_ε by $\pi_\varepsilon[\mathbf{i}](\mathbf{j}, u) := \pi_\varepsilon(\mathbf{i}+\mathbf{j}, u)$.

   The following claim then implies that $\mathcal{R}^\sigma(\mathbf{i}, v) := \lim_{k\to\infty} \mathcal{R}^\sigma_k(\mathbf{i}, v) \leq \mu\mathcal{V}(\mathbf{i}, v) + \varepsilon$.

   Claim 3.8. For every $\sigma \in \Sigma$, $k \geq 0$, $\mathbf{i} \in \mathcal{H}$, $v \in V$, $\varepsilon \geq 0$, we have

$$\mathcal{R}^\sigma_k(\mathbf{i}, v) \leq \mu\mathcal{V}(\mathbf{i}, v) + \sum_{j=1}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}}$$

Both Claim 3.7 and 3.8 are purely technical, for proofs see Appendix C.1. □

It follows directly from Definition 2.3 and Theorem 3.6 that both players have ε-optimal strategies in every vertex v (for every $\varepsilon > 0$). Now we examine the existence of *optimal* strategies. We start by observing that optimal strategies do not necessarily exist in general.

Observation 3.9. *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 $V = \{v_i \mid i \in \mathbb{N}_0\} \cup \{start, down\}$, $A = \{a_i, b_i \mid i \in \mathbb{N}\} \cup \{c, d\}$, $\mathbf{E}(start) = \{a_i \mid i \in \mathbb{N}\}$, $\mathbf{E}(v_i) = \{b_i\}$ for all $i \in \mathbb{N}$, $\mathbf{E}(v_0) = \{c\}$, $\mathbf{E}(down) = \{d\}$, $\mathbf{P}(a_i)(v_i) = 1$, $\mathbf{P}(c)(v_0) = 1$, $\mathbf{P}(d)(down) = 1$, 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 in Figure 1 (the partition of V into $(V_\Box, V_\Diamond)$ is not fixed yet, and the vertices are therefore drawn as ovals). If we put $V_\Box = V$, we obtain that $\sup_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi}_{start}(Reach^{\leq 1}(\{down\})) = \sum_{\ell=1}^{\infty} \left(\frac{1}{2^\ell} F_{\ell+1}(1)\right)$ where π

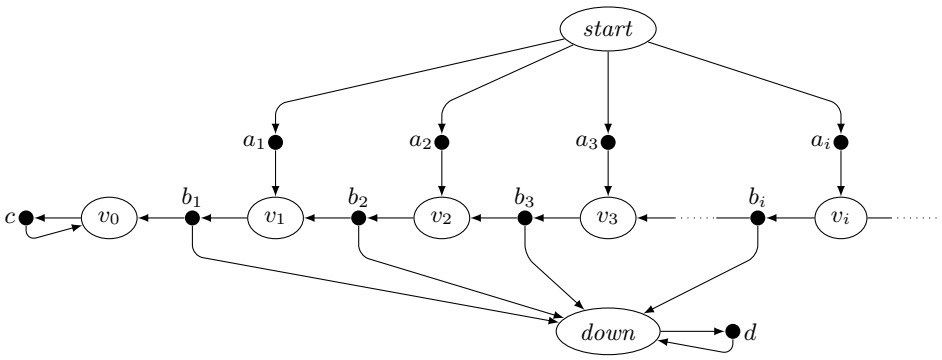


Figure 1: Optimal strategies may not exist.

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 $\inf_{\pi \in \Pi} \mathcal{P}^{\sigma,\pi}_{start}(\mathit{Reach}^{\leq 1}(\{v_0\})) = 0$ where σ is the trivial strategy for player $\Box$, but there is no optimal minimizing strategy. $\Box$

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 3.6.

Theorem 3.10. *If G is finitely-branching, then there is an optimal minimizing CD strategy.*

Proof. It suffices to reconsider the second assertion of the proof of Theorem 3.6. Since G is finitely-branching, the infima over enabled actions in the definition of $\mathcal{V}$ are actually minima. Hence, in the definition of π_ε, we can set $\varepsilon = 0$ and pick actions yielding minimal values. Thus the strategy π_ε becomes an optimal minimizing CD strategy. $\Box$

Observe that for optimal minimizing strategies we did not require that G has bounded rates. 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 3.11. *Optimal maximizing strategies in continuous-time games with time-bounded reachability objectives may not exist, even if we restrict ourselves to finitely-branching games.*

Proof. Consider a game $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$, where $V = V_\Box = \{v_i, u_i \mid i \in \mathbb{N}_0\} \cup \{\mathit{win}, \mathit{lose}\}$; $A = \{a_i, b_i, \mathit{end}_i \mid i \in \mathbb{N}_0\} \cup \{w, \ell\}$, $\mathbf{E}(\mathit{win}) = \{w\}$, $\mathbf{E}(\mathit{lose}) = \{\ell\}$, and $\mathbf{E}(v_i) = \{a_i, b_i\}$, $\mathbf{E}(u_i) = \{\mathit{end}_i\}$ for all $i \in \mathbb{N}_0$; $\mathbf{R}(w) = \mathbf{R}(\ell) = 1$, and $\mathbf{R}(a_i) = \mathbf{R}(b_i) = 2^i$, $\mathbf{R}(\mathit{end}_i) = 2^{i+1}$ for all $i \in \mathbb{N}_0$; $\mathbf{P}(w)(\mathit{win}) = \mathbf{P}(\ell)(\mathit{lose}) = 1$, and for all $i \in \mathbb{N}_0$ we have that $\mathbf{P}(a_i)(v_{i+1}) = 1$, $\mathbf{P}(b_i)(u_i) = 1$,

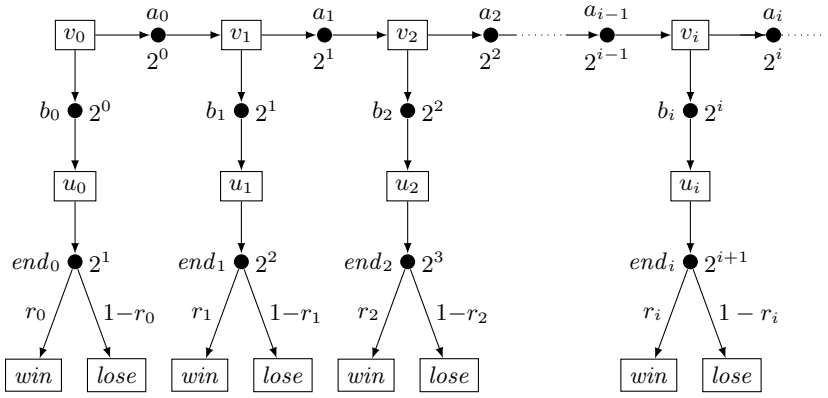


Figure 2: Optimal maximizing strategies may not exist in finitely-branching games.

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 in Figure 2 (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 $q > \frac{1}{2}$ such that $F_{\mathbf{i}_k}(t) \geq q$ for every $k \in \mathbb{N}$. Note that such t and q exist because the mean value associated to $F_{\mathbf{i}_k}$ is $\sum_{i=0}^{k} 1/2^i < 2$ and hence it suffices to apply Markov inequality. 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 $\sup_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi}_{v_0}(\mathit{Reach}^{\leq t}(\{\mathit{win}\})) = q$ (where π is the trivial strategy for player $\Diamond$), but there is no strategy σ such that $\mathcal{P}^{\sigma,\pi}_{v_0}(\mathit{Reach}^{\leq t}(\{\mathit{win}\})) = q$. The first part follows by observing that player $\Box$ can reach *win* within time t 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 less than q, and by not selecting b_j at all, this probability becomes equal to 0. $\square$

Observe that again the counterexample is a CTMDP. Now we show that if G is finitely-branching *and* has bounded rates, then there is an optimal maximizing CD strategy. First, observe that for each $k \in \mathbb{N}_0$

$$\sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathcal{P}^{\sigma,\pi}_{v}(\mathit{Reach}^{\leq t}_{\leq k}(T)) \;=\; \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \mathcal{P}^{\sigma,\pi}_{v}(\mathit{Reach}^{\leq t}_{\leq k}(T)) \;=\; \mathcal{V}^{k+1}(\mathit{zero})(\mathbf{0}, v) \tag{1}$$

where $\mathcal{V}$ is the function defined in the proof of Theorem 3.6, $\mathit{zero} : \mathcal{H} \times V \to [0,1]$ is a constant function returning zero for every argument, and $\mathbf{0}$ is the empty history. A proof of Equality 1 is obtained by a straightforward induction on k. We use $\mathit{val}^k(v)$ to denote the

k-step value defined by Equality 1, 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 $\inf_{\pi \in \Pi} \mathcal{P}_v^{\sigma_k,\pi}(Reach_{\leq k}^{\leq t}(T)) = \sup_{\sigma \in \Sigma} \mathcal{P}_v^{\sigma,\pi_k}(Reach_{\leq k}^{\leq t}(T)) = val^k(v)$. The existence and basic properties of k-step optimal strategies are stated in our next lemma.

Lemma 3.12. *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*

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

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 ε-optimal strategy.*

Proof. See Appendix C.2. □

If G is finitely-branching and has bounded rates, one may be tempted to construct an optimal maximizing strategy σ 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 3.12 (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 3.13. *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 $R \subseteq Run_{G(\hat{v},\sigma,\pi)}$, we denote $\mathcal{P}_{\hat{v}}^{\sigma,\pi}(R)$ the probability of R in $G(\hat{v}, \sigma, \pi)$. For every $k \in \mathbb{N}$ we fix a k-step optimal BCD strategy σ_k of player $\square$ (see Lemma 3.12). Let us order the set $\mathcal{R}$ of rates into an enumerable sequence $r_1, r_2 \ldots$ and the set V into an enumerable sequence $v_1, v_2 \ldots$. 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 Γ_ℓ to be an infinite subset of $\Gamma_{\ell-1}$ such that we have $\sigma(\mathbf{i}, v_n) = \sigma'(\mathbf{i}, v_n)$ 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 σ. Let $\mathbf{i} \in \mathcal{H}$ and $v_n \in V$, we choose a number ℓ such that $\ell > |\mathbf{i}|$, $\ell > n$ and $\mathbf{i}(j) = 0$ for all $j > \ell$ (note that such ℓ exists for any $\mathbf{i} \in \mathcal{H}$ and $v_n \in V$). We put $\sigma(\mathbf{i}, v_n) = \sigma'(\mathbf{i}, v_n)$ where $\sigma' \in \Gamma_\ell$. It is easy to see that σ is a CD strategy, it remains to argue that it is optimal. Suppose the converse, i.e. that it is not ε-optimal in some v_{in} for some $\varepsilon > 0$.

Let us fix k satisfying conditions of part 1 of Lemma 3.12 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 σ' and π' and all k we have that $\mathcal{P}_v^{\sigma',\pi'}(U_i^{v,\sigma',\pi'}) \leq \frac{\varepsilon}{2k}$ where $U_i^{v,\sigma',\pi'}$ is the set of all runs of $G(v,\sigma',\pi')$ that *do not* contain any state of the form $v_0 a_0 \dots a_{i-1} v_i a_i$ where $v_i \in B_{a_{i-1}}$. As a consequence we have $\mathcal{P}_v^{\sigma',\pi'}(\bigcap_{i=0}^{k} U_i^{v,\sigma',\pi'}) \leq \frac{\varepsilon}{4}$. In the sequel, we denote $U^{v,\sigma',\pi'} = \bigcap_{i=0}^{k} U_i^{v,\sigma',\pi'}$ and we write just U instead of $U^{v,\sigma',\pi'}$ if v, σ and π are clear from the context.

Let W be the set of histories of the form $v_0 a_0 \dots v_{i-1} a_{i-1} v_i$ where $i \leq k$, $v_0 = v_{in}$, and for all $0 \leq j < i$ we have $a_j \in \mathbf{E}(v_j)$, $\mathbf{P}(a_j)(v_{i+j}) > 0$ and $v_{j+1} \notin B_{a_j}$. We claim that there is $m \geq n$ s.t. σ_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 ℓ such that $k \leq \ell$ and for all $w \in W$, there is no v_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 π.

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

Finally, we obtain

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

which means that σ is ε-optimal in v_{in}. □

4. Optimal Strategies in Finite Uniform CTGs

In this section, we restrict ourselves to finite uniform CTGs, i.e. $\mathbf{R}(a) = r > 0$ for all $a \in A$. The histories from $\mathcal{H}$ are thus vectors of length 1, hence we write them as integers. We prove that both players have *optimal BCD strategies* in such games. More precisely, we prove a stronger statement that there are optimal strategies that after some number of steps eventually behave in a stationary way. A CD strategy τ is *stationary* if $\tau(h, v)$ depends just on v for every vertex v. Besides, for a CD strategy τ, a strategy $\tau[h]$ is defined by $\tau[h](h', v) = \tau(h + h', v)$. Further, recall that *bp* is the maximum bit-length of the fractional representation of transition probabilities.

Theorem 4.14. *In a finite uniform CTG, there exist optimal CD strategies $\sigma \in \Sigma$, $\pi \in \Pi$ and $k \in \mathbb{N}$ such that $\sigma[k]$ and $\pi[k]$ are stationary; in particular, σ*

and π are optimal BCD strategies. Moreover, if all transition probabilities are rational then one can choose $k = rt(1 + 2^{bp \cdot |A|^2 \cdot |V|^3})$.

We then 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 even if the branching is finite (see Observation 4.23). In Section 5, we use these results to design an algorithm which *computes* the optimal BCD strategies in finite uniform games. Further, using the method of uniformization where a general game is reduced to a uniform one, the results can be extended to general (not necessarilly uniform) finite games, see [**?**].

Before proving the theorem we note that the crucial point is to understand the behaviour of optimal strategies after many (i.e. k) steps have already been taken. In such a situation, not much time is left and it turns out that in such a situation optimal strategies optimize the probability of reaching T in as few steps as possible. This motivates the central definition of *greedy* strategies. Intuitively, a greedy strategy optimizes the outcome of the first step. If there are more options to do so, it chooses among these options so that it optimizes the second step, etc.

Definition 4.15. For strategies $\sigma \in \Sigma$ and $\pi \in \Pi$ and a vertex v, we define a *step reachability vector* $\overrightarrow{\mathcal{P}}_v^{\sigma,\pi} = \left(\mathcal{P}_v^{\sigma,\pi}(Reach_{=i}^{\leq\infty}(T))\right)_{i \in \mathbb{N}_0}$. A strategy $\sigma \in \Sigma$ is *greedy* if for every v, $\min_{\pi \in \Pi} \overrightarrow{\mathcal{P}}_v^{\sigma,\pi} = \max_{\sigma' \in \Sigma} \min_{\pi \in \Pi} \overrightarrow{\mathcal{P}}_v^{\sigma',\pi}$ where the optima[3] are considered in the lexicographical order. Similarly, a strategy $\pi \in \Pi$ is *greedy* if $\max_{\sigma \in \Sigma} \overrightarrow{\mathcal{P}}_v^{\sigma,\pi} = \min_{\pi' \in \Pi} \max_{\sigma \in \Sigma} \overrightarrow{\mathcal{P}}_v^{\sigma,\pi'}$ for every v.

We prove the theorem as follows:

1. Optimal CD strategies are guaranteed to exist by Theorem 3.10 and Theorem 3.13.

2. For every optimal CD strategy τ, the strategy $\tau[k]$ is *greedy* (see Proposition 4.16).

3. There exist *stationary* greedy strategies (see Proposition 4.21). Let τ_g be such a strategy. Then for an optimal strategy τ, the strategy $\bar{\tau}$ defined by

$$\bar{\tau}(h, v) = \begin{cases} \tau(h, v) & \text{if } h < k; \\ \tau_g(h, v) & \text{otherwise} \end{cases}$$

   is clearly BCD and also optimal. Indeed, all greedy strategies guarantee the same probabilities to reach the target. (This is clear by definition, since their step reachability vectors are the same.) Therefore, we can freely interchange them without affecting the guaranteed outcome.

[3]We can use optima instead of extrema as the optimal strategies obviously exist in finite discrete-time (with the time bound being infinite) games even when the number of steps is fixed.

Proposition 4.16. *Let τ be an optimal strategy. Then there is $k \in \mathbb{N}$ such that $\tau[k]$ is greedy. Moreover, if all transition probabilities are rational then one can choose $k = rt(1 + 2^{bp\cdot|A|^2\cdot|V|^3})$.*

In order to prove the proposition, we relax our definition of greedy strategies. A strategy is *greedy on s steps* if the greedy strategy condition holds for the step reachability vector where only first s elements are considered. A strategy τ is *always greedy on s steps* if for all $i \in \mathbb{N}_0$ the strategy $\tau[i]$ is greedy on s steps. We use this relaxation of greediness to prove the proposition as follows. We firstly prove that every optimal strategy is always greedy on $|\mathbf{E}| := \sum_{v\in V} |\mathbf{E}(v)|$ steps (by instantiating Lemma 4.17 for $s = |\mathbf{E}| \leq |A| \cdot |V|$) and then Lemma 4.18 concludes by proving that being always greedy on $|\mathbf{E}|$ steps guarantees greediness.

Lemma 4.17. *For every $s \in \mathbb{N}$ there is $\delta > 0$ such that for every optimal CD strategy τ the strategy $\tau[rt(1 + 1/\delta)]$ is always greedy on s steps. Moreover, if all transition probabilities are rational, then one can choose $\delta = 1/2^{bp\cdot|V|\cdot|\mathbf{E}|\cdot s}$.*

Proof. We look for a δ such that for every optimal strategy $\sigma \in \Sigma$ if $\sigma[h]$ is not greedy on s steps then $h < k$, where $k = rt(1 + 1/\delta)$. Let thus σ be an optimal CD strategy and $s \in \mathbb{N}$. For $\sigma[h]$ that is not greedy on s steps there is $i \leq s$, a vertex v and a strategy σ^* such that

$$\big(\inf_{\pi\in\Pi} \overrightarrow{\mathcal{P}}_v^{\sigma[h],\pi}\big)_i < \big(\inf_{\pi\in\Pi} \overrightarrow{\mathcal{P}}_v^{\sigma^*,\pi}\big)_i$$

and for all $j < i$

$$\big(\inf_{\pi\in\Pi} \overrightarrow{\mathcal{P}}_v^{\sigma[h],\pi}\big)_j = \big(\inf_{\pi\in\Pi} \overrightarrow{\mathcal{P}}_v^{\sigma^*,\pi}\big)_j$$

This implies that there is $i \leq s$ such that $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma^*,\pi}(Reach_{\leq i}^{\leq\infty}(T)) - \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma[h],\pi}(Reach_{\leq i}^{\leq\infty}(T))$ is positive. Since the game is finite there is a fixed $\delta > 0$ such that difference of this form is (whenever it is non-zero) greater than δ for all deterministic strategies σ and σ^*, $v \in V$ and $i \leq s$. Moreover, if all transition probabilities are rational, then δ can be chosen to be $1/M^s$, where M is the least common multiple of all probabilities denominators. Indeed, $\mathcal{P}_v^{\sigma,\tau}(Reach_{\leq i}^{\leq\infty}(T))$ is clearly expressible as ℓ/M^i for some $\ell \in \mathbb{N}_0$. Since there are at most $|V| \cdot |\mathbf{E}|$ probabilities, we have $\delta \geq 1/2^{bp\cdot|V|\cdot|\mathbf{E}|\cdot s}$.

We define a (not necessarily counting) strategy $\bar{\sigma}$ that behaves like σ, but when h steps have been taken and v is reached, it behaves as σ^*. We show that for $h \geq k$ this strategy $\bar{\sigma}$ would be an improvement against the optimal strategy σ. There is clearly an improvement at the $h + i$th step provided one gets there on time, and this improvement is at least δ. Nonetheless, in the next steps there may be an arbitrary decline. Altogether due to optimality of σ

$$\begin{aligned}
&0 \geq \inf_{\pi\in\Pi} \mathcal{P}_{\hat{v}}^{\bar{\sigma},\pi}(Reach^{\leq t}(T)) - \inf_{\pi\in\Pi} \mathcal{P}_{\hat{v}}^{\sigma,\pi}(Reach^{\leq t}(T)) \geq \\
&\geq \inf_{\pi\in\Pi} \mathcal{P}_{\hat{v}}^{\sigma,\pi}(\overset{h}{\to} v) \cdot \Big[F_{h+i}(t) \cdot \delta - F_{h+i+1}(t) \cdot 1\Big] = (*)
\end{aligned}$$

where $\mathcal{P}^{\sigma,\pi}_{\hat{v}}(\xrightarrow{h} v)$ is the probability that after h steps we will be in v. We need to show that the inequality $0 \geq (*)$ implies $h < k$. We use the following key argument that after taking sufficiently many steps, the probability of taking strictly *more* than one step before the time limit is negligible compared to the probability of taking *precisely* one more step, i.e. that for all $n \geq k = rt(1+1/\delta)$ we have

$$\frac{F_{n+1}(t)}{F_n(t)} < \frac{F_{n+1}(t)}{F_n(t) - F_{n+1}(t)} < \delta$$

As $F_{n+1}(t) = \sum_{i=1}^{\infty} e^{-r\cdot t}(rt)^{n+i}/(n+i)!$, this is proved by the following:

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

This argument thus implies $h + i < k$, hence we conclude that indeed $h < k$. The minimizing part is dual. □

The following lemma concludes the proof of the proposition.

Lemma 4.18. *A strategy is greedy iff it is always greedy on $|\mathbf{E}|$ steps.*

Proof. We need to focus on the structure of greedy strategies. Therefore, we provide their inductive characterization. Moreover, this characterization can be easily turned into an algorithm computing all greedy strategies.

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.

Algorithm 1 computes which actions can be chosen in greedy strategies. We begin with the original game and keep on pruning inoptimal transitions until we reach a fix-point. In the first iteration, we compute the value $R_1(v)$ for each vertex v, which is the optimal probability of reaching T in one step. We remove all transitions that are not optimal in this sense. In the next iteration, 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 only the remaining "greedy" transitions thus results in greedy behavior.

Claim 4.19. A strategy is always greedy on s steps iff it uses transitions from $\mathbf{E}_s$ only (as defined by Algorithm 1).

In particular, a strategy τ is always greedy on $|\mathbf{E}|$ steps iff it uses transitions from $\mathbf{E}_{|\mathbf{E}|}$ only. For the proof of Claim 4.19 see Appendix C.3.

Claim 4.20. A strategy is greedy iff it uses transitions from $\mathbf{E}_{|\mathbf{E}|}$ only.

The proof of Claim 4.20 now follows easily. Since the number of edges is finite, there is a fix-point $\mathbf{E}_n = \mathbf{E}_{n+1}$, moreover, $n \leq |\mathbf{E}|$. Therefore, any strategy using $\mathbf{E}_{|\mathbf{E}|}$ only is by Claim 4.19 always greedy on s steps for all $s \in \mathbb{N}_0$, hence clearly greedy. On the other hand, every greedy strategy is in particular always greedy on $|\mathbf{E}|$ steps and thus uses transitions from $\mathbf{E}_{|\mathbf{E}|}$ only again by Claim 4.19. This concludes the proof of the Lemma and thus also of Proposition 4.16. □

Algorithm 1 computing all greedy edges

$$R_0(v) = \begin{cases} 1 & \text{if } v \in T, \\ 0 & \text{otherwise.} \end{cases}$$
$$\mathbf{E}_0(v) = \mathbf{E}(v)$$

$$R_{i+1}(a) = \sum_{u \in V} \mathbf{P}(a)(u) \cdot R_i(u)$$
$$R_{i+1}(v) = \begin{cases} \max_{a \in \mathbf{E}_i(v)} R_{i+1}(a) & \text{if } v \in V_\Box, \\ \min_{a \in \mathbf{E}_i(v)} R_{i+1}(a) & \text{otherwise.} \end{cases}$$
$$\mathbf{E}_{i+1}(v) = \mathbf{E}_i(v) \cap \{a \mid R_{i+1}(a) = R_{i+1}(v)\}$$

We now move on to Proposition 4.21 that concludes the proof of the theorem.

Proposition 4.21. *There are greedy stationary strategies $\sigma_g \in \Sigma$ and $\pi_g \in \Pi$. Moreover, the strategies σ_g and π_g are computable in polynomial time.*

Proof. The complexity of Algorithm 1 is polynomial in the size of the game graph as the fix-point is reached within $|\mathbf{E}|$ steps. And 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 strategy (by Claim 4.20) that is stationary. □

Corollary 4.22. *In a finite uniform game with rational transition probabilities, there are optimal strategies τ such that $\tau[rt(1 + 2^{bp \cdot |A|^2 \cdot |V|^3})]$ is a greedy stationary strategy.*

A natural question is whether Theorem 4.14 and Corollary 4.22 can be extended to infinite-state uniform CTGs. The question is answered in our next observation.

Observation 4.23. *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.*

Proof. Consider a game $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$ where $V = V_\Box = \{v_i, u_i, \bar{u}_i, \hat{u}_i \mid i \in \mathbb{N}_0\} \cup \{down\}$, $A = \{a_i, hat_i, bar_i, \hat{b}_i, \bar{b}_i \mid i \in \mathbb{N}_0\}$, $\mathbf{E}(v_i) = \{a_i\}$, $\mathbf{E}(u_i) = \{bar_i, hat_i\}$, $\mathbf{E}(\hat{u}_i) = \{\hat{b}_i\}$, 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 $\{v_0, v_1, u_0\}$, $\mathbf{P}(a_i)$ is the uniform distribution on $\{u_i, v_{i+1}\}$ for $i > 0$,
- $\mathbf{P}(hat_i)(\hat{u}_i) = 1$ and $\mathbf{P}(bar_i)(\bar{u}_i) = 1$ for $i \geq 0$,

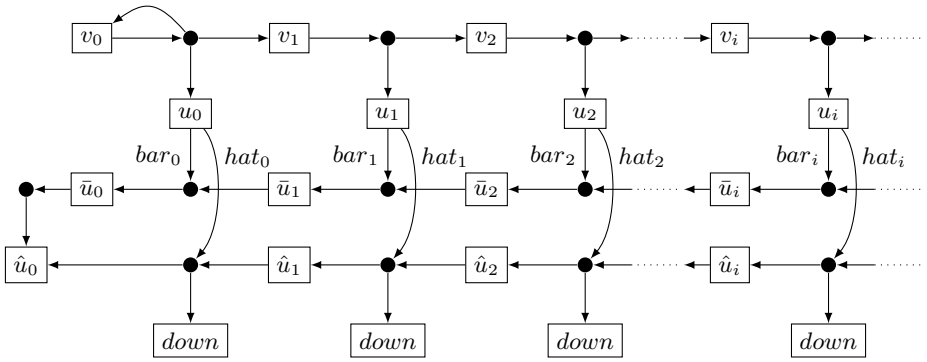


Figure 3: Optimal BCD strategies may not exist in infinite uniform games.

- $\mathbf{P}(\bar{b}_0)(\hat{u}_0) = 1$, and $\mathbf{P}(\bar{b}_i)(\bar{u}_{i-1}) = 1$ for $i > 0$,
- $\mathbf{P}(\hat{b}_i)$ is the uniform distribution on $\{\hat{u}_{i-1}, down\}$ for $i \geq 1$.

We set $\mathbf{R}(a) = 1$ for all $a \in A$. The structure of G is shown in Figure 3. Observe that player $\Box$ has a real choice only in states u_i.

We show that if the history is of the form $v_0a_0v_1a_1 \dots v_ia_iu_i$ (where $i \in \mathbb{N}_0$), the optimal strategy w.r.t. reaching $\hat{u}_0$ within time $t = 1$ 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 $\frac{F_{2i+2}(1)}{F_{2i+3}(1)} < 2^{i+2}$, for infinitely many is. This follows by observing that for $i > 0$

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

On the other hand, from Lemma 4.17 one can deduce that for all i there is $j \geq i$ such that any optimal strategy must choose hat_i if the history is of the form $(v_0a_0)^j v_1a_1 \dots v_ia_iu_i$. Thus no strategy with counting bounded by $k \in \mathbb{N}$ can be optimal as one can choose $j \geq i > k$.

□

5. Algorithms

Now we present algorithms which compute ε-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*.

5.1. Computing ε-optimal BCD strategies

For this subsection, let us fix a CTG $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$, a set $T \subseteq V$ of target vertices, a time bound $t > 0$, and some $\varepsilon > 0$. For simplicity, let us

Algorithm 2 Compute the function C

{1st phase: compute the approximations of $F_{\mathbf{i}}(t)$ and $\mathbf{P}$}
for all vectors $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$ **do**
  compute a number $\ell_{\mathbf{i}}(t) > 0$ such that $\frac{|F_{\mathbf{i}}(t) - \ell_{\mathbf{i}}(t)|}{F_{\mathbf{i}}(t)} \leq \left(\frac{\varepsilon}{2}\right)^{2|\mathbf{i}|+1}$.
for all actions $a \in A$ and vertices $u \in V$ **do**
  compute a floating point representation $\mathbf{p}(a)(u)$ of $\mathbf{P}(a)(u)$ satisfying $\frac{|\mathbf{P}(a)(u) - \mathbf{p}(a)(u)|}{\mathbf{P}(a)(u)} \leq \left(\frac{\varepsilon}{2}\right)^{2k+1}$.
{2nd phase: compute the functions R and C in a bottom up manner}
for all vector lenghts j from k down to 0 **do**
  for all vectors $\mathbf{i} \in \mathcal{H}$ of length $|\mathbf{i}| = j$ **do**
    for all vertices $v \in V$ **do**
      if $v \in T$ **then**
        $R(\mathbf{i}, v) \leftarrow \ell_{\mathbf{i}}(t)$
      else if $|\mathbf{i}| = k$ **then**
        $R(\mathbf{i}, v) \leftarrow 0$
      else if $v \in V_{\Box}$ **then**
        $R(\mathbf{i}, v) \leftarrow \max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u)$
        $C(\mathbf{i}, v) \leftarrow a$ where a is the action that realizes the maximum above
      else if $v \in V_{\Diamond}$ **then**
        $R(\mathbf{i}, v) \leftarrow \min_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u)$
        $C(\mathbf{i}, v) \leftarrow a$ where a is the action that realizes the minimum above

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 5.26). Let $k = (\max \mathcal{R})te^2 - \ln(\frac{\varepsilon}{2})$. Then, due to Lemma 3.12, all k-step optimal strategies are $\frac{\varepsilon}{2}$-optimal. We use the remaining $\frac{\varepsilon}{2}$ for numerical imprecisions.

We need to specify the ε-optimal BCD strategies $\sigma_\varepsilon \in \Sigma$ and $\pi_\varepsilon \in \Pi$ on the first k steps. For every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| < k$, 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 strategies. That is, for every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| < k$, and for every $v \in V_{\Box}$, we put $\sigma_\varepsilon(\mathbf{i}, v)(C(\mathbf{i}, v)) = 1$, and for the other arguments we define σ_ε arbitrarily so that σ_ε remains a BCD strategy. The strategy π_ε is induced by the function C in the same way.

The procedure to compute the function C is described in Algorithm 2. For computing $C(\mathbf{i}, v)$ it 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 σ_ε and π_ε (precisely, using the parts of strategies σ_ε and π_ε computed so far) 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 σ_ε and π_ε are ε-optimal, but they are not necessarily k-step optimal for the k chosen above).

Lemma 5.24. *The strategies σ_ε and π_ε are ε-optimal.*

Proof. See Appendix C.4. □

Assuming that the probabilities $\mathbf{P}(a)(u)$ and rates are given as fractions with both numerator and denominator represented in binary with length bounded by bp, a complexity analysis of the algorithm reveals the following.

Theorem 5.25. *Assume that G is finite. Then for every $\varepsilon > 0$ there are ε-optimal BCD strategies $\sigma_\varepsilon \in \Sigma$ and $\pi_\varepsilon \in \Pi$ computable in time $|V|^2 \cdot |A| \cdot bp^2 \cdot \left((\max \mathcal{R}) \cdot t + \ln \frac{1}{\varepsilon}\right)^{2|\mathcal{R}|+\mathcal{O}(1)}$.*

Proof. We analyze the complexity of Algorithm 2. We start with 1st phase. Recall that $k = (\max \mathcal{R})te^2 + \ln \frac{1}{\varepsilon}$. (Here we use ε instead of $\varepsilon/2$ as this difference is clearly absorbed in the $\mathcal{O}$-notation.)

We approximate the value of $F_{\mathbf{i}}(t)$ to the relative precision $(\varepsilon/2)^{2k+1}$ as follows. According to [**?**], the value of $F_{\mathbf{i}}(t)$ is expressible as $\sum_{r \in \mathcal{R}} q_r e^{-rt}$. First, q_r here is a polynomial in t and can be precisely computed using polynomially many (in $|\mathbf{i}| \leq k$ and $|\mathcal{R}|$) arithmetical operations on integers with length bounded by $bp + \ln k + \ln t$. Hence the computation of q_r as a fraction can be done in time $bp^2 \cdot k^{\mathcal{O}(1)} \cdot |\mathcal{R}|^{\mathcal{O}(1)}$ and both the numerator and the denominator are of length $bp \cdot k^{\mathcal{O}(1)} \cdot |\mathcal{R}|^{\mathcal{O}(1)}$. 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 the fraction and $k \ln \frac{1}{\varepsilon}$, hence again in the time $bp^2 \cdot k^{\mathcal{O}(1)} \cdot |\mathcal{R}|^{\mathcal{O}(1)}$. Secondly, according to [**?**], the floating point approximation of e^{-rt} with the relative error $(\varepsilon/4)^{2k+1}$ can be computed in time less than quadratic in $k \ln \frac{1}{\varepsilon}$. Altogether, we can compute an $(\varepsilon/2)^{2k+1}$-approximation of each $F_{\mathbf{i}}(t)$ in time $|\mathcal{R}| \cdot bp^2 \cdot k^{\mathcal{O}(1)} \cdot |\mathcal{R}|^{\mathcal{O}(1)} = bp^2 \cdot k^{\mathcal{O}(1)} \cdot |\mathcal{R}|^{\mathcal{O}(1)}$. This procedure has to be repeated for every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$. The number of such $\mathbf{i}$'s is bounded by $\binom{|\mathcal{R}|+k}{k} \leq \mathcal{O}(k^{|\mathcal{R}|})$. So computing all $(\varepsilon/2)^{2k+1}$-approximations $\ell_{\mathbf{i}}(t)$ of values $F_{\mathbf{i}}(t)$ can be done in time $\mathcal{O}(k^{|\mathcal{R}|}) \cdot bp^2 \cdot k^{\mathcal{O}(1)} \cdot |\mathcal{R}|^{\mathcal{O}(1)} \subseteq bp^2 \cdot k^{|\mathcal{R}|+\mathcal{O}(1)}$.

Using a 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 $(\varepsilon/2)^{2k+1}$ in time linear in $bp \cdot k \ln \frac{1}{\varepsilon}$. So the first phase takes time $bp^2 \cdot k^{|\mathcal{R}|+\mathcal{O}(1)} + |A| \cdot |V| \cdot \mathcal{O}(bp \cdot k \ln \frac{1}{\varepsilon}) \subseteq |V| \cdot |A| \cdot bp^2 \cdot k^{|\mathcal{R}|+\mathcal{O}(1)}$.

In 2nd phase, 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 $\binom{|\mathcal{R}|+k}{k} \cdot |V| \leq \mathcal{O}(k^{|\mathcal{R}|} \cdot |V|)$.

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

Altogether, the overall time complexity of Algorithm 2 is bounded by

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

□

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 5.26. *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 ε-optimal in v.*

Proof. By Lemma 3.12, 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 $v_0 \dots v_k \in V^*$ and $a_0 \dots a_k \in A^*$ such that, $v_0 = v$, $v_k = v'$, $a_i \in \mathbf{E}(v_i)$ for all $0 \leq i \leq k$ and $\mathbf{P}(a_i)(v_{i+1}) > 0$ for all $0 \leq i < k$). 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 5.25 there is an $\frac{\varepsilon}{4}$-optimal BCD strategy σ' in this game. Now it suffices to extend σ' to a BCD strategy σ in the original game by defining, arbitrarily, its values for vertices and actions removed by the above procedure. It is easy to see that σ is an ε-optimal BCD strategy in G. □

5.2. 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 $k = rt(1 + 2^{bp \cdot |A|^2 \cdot |V|^3})$ (see Corollary 4.22).

The algorithm works similarly as the one of Section 5.1, 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 5.1, 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 Corollary 4.22, 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 σ_{max} and π_{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 σ_{max} and π_{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

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

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 5.27. *The BCD strategies σ_{max} and π_{min} are optimal and effectively computable.*

Proof. We start by showing that σ_{max} and π_{min} are optimal. Let us denote by Σ_g (resp. Π_g) the set of all CD strategies $\sigma \in \Sigma$ (resp. $\pi \in \Pi$) such that for all $u \in V_\Box$ ($u \in V_\Diamond$) and $i \geq k$ we have $\sigma(i,u) = \sigma_g(u)$, which is a stationary greedy strategy. By Corollary 4.22, for every $v \in V$ we have

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

Recall that given a CD strategy τ and $i \geq 0$, we denote by $\tau[i]$ a strategy obtained from τ by $\tau[i](j,u) = \tau(i+j,u)$. Let us denote

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

For every $i \geq 0$ we put

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

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

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

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

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

Using this fomulation, the optimality of σ_{max} and π_{min} is proven in the following claim.

Claim 5.28. Let $i \leq k$ and $v \in V$. We have

$$\min_{\pi \in \Pi_g} \bar{K}^{\pi}(i,v) = R(i,v) = \max_{\sigma \in \Sigma_g} \bar{K}^{\sigma}(i,v) \tag{2}$$

$$R(i,v) = val(i,v) \tag{3}$$

Proof. We start by proving the equation (2). If $v \in T$, then $\bar{K}^{\pi}(i,v) = \bar{K}^{\sigma}(i,v) = F_i(t) = R(i,v)$. 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 claim holds true for n and consider $n+1$. If $v \in V_{\Box}$ and $\sigma_{max}(\mathbf{i},v)(b) = 1$,

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

and

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

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

Now the equation (3) follows easily:

$$\begin{aligned}
R(i,v) = \min_{\pi \in \Pi_g} \bar{K}^{\pi}(i,v) \leq \max_{\sigma \in \Sigma_g} \min_{\pi \in \Pi_g} \bar{P}^{\sigma,\pi}(i,v) = \\
\min_{\pi \in \Pi_g} \max_{\sigma \in \Sigma_g} \bar{P}^{\sigma,\pi}(i,v) \leq \max_{\sigma \in \Sigma_g} \bar{K}^{\sigma}(i,v) = R(i,v)
\end{aligned}$$

□

This proves that σ_{max} and π_{min} are optimal.

Effective computability of σ_{max} and π_{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 $\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma_g,\pi_g}(Reach_{=j}^{<\infty}(T))$ can effectively be expressed as a linear combination of transcendental numbers of the form e^{ct} where c is algebraic. Therefore, each difference of the compared numbers can effectively be expressed as a finite sum $\sum_j \eta_j e^{\delta_j}$ where the η_j and δ_j are algebraic numbers and the δ_j's are pairwise distinct. Now it suffices to apply Lemma 2 of [**?**] to decide whether the difference is greater than 0, or not.

It remains to show that $\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma_g,\pi_g}(Reach_{=j}^{<\infty}(T))$ 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 $1 \leq j \leq i$, and setting $\mathbf{P}(a_i)(v) = 1$, and $\mathbf{P}(a_j)(v_{j+1}) = 1$ for $1 \leq j < i$ (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 $1 \leq j \leq i$. As the strategies σ_g and π_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 $1 \leq j \leq i$ we obtain

$$\mathcal{P}_{v_1}^{\sigma_g,\pi_g}(Reach^{\leq t}(T)) = \sum_{j=0}^{\infty} F_j(t) \cdot \mathcal{P}_{v_1}^{\sigma_g,\pi_g}(Reach_{=j}^{<\infty}(T)) =$$

$$\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_{v_1}^{\sigma_g,\pi_g}(Reach_{=i+j}^{<\infty}(T)) = \sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma_g,\pi_g}(Reach_{=j}^{<\infty}(T))$$

As σ_g and π_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 [**?**] and obtain the desired form of $\mathcal{P}_{v_1}^{\sigma_g,\pi_g}(Reach^{\leq t}(T))$, and hence also of $\sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma_g,\pi_g}(Reach_{=j}^{<\infty}(T))$. □

6. 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 ε-optimal BCD strategies in finitely-branching CTGs with bounded rates, and optimal BCD strategies in finite uniform CTGs.

There are at least two 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 [**?**], it is demonstrated that time-dependent strategies are more powerful (i.e., can achieve better results) than the time-abstract ones. However, this issue is somewhat subtle—in [**?**], it is shown that the power of time-dependent strategies is different when the player knows only the total elapsed time, the time consumed by the last action, or the complete timed history of a play. The analog of Theorem 3.6 in this setting is

examined in [**?**]. In [**?**] ε-optimal time-dependent strategies are computed for CTMDPs. Second, a generalization to semi-Markov processes and games, where arbitrary (not only exponential) distributions are considered, would be desirable.

Appendix 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* λ 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.29. *Let $\mathcal{M} = (S, \mathbf{P}, \mathbf{R}, \mu)$ be a CTMC, $j \in \mathbb{N}_0$, $t \in \mathbb{R}^{\geq 0}$, and $u_0, \ldots, u_j \in S$. Let U be the set of all runs (u, s) where u starts with $u_0, \ldots, u_j$ and $\sum_{i=0}^{j} s_j \leq t$. We have that*

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

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 $\mathcal{P}(U) = \mu(u_0)$ 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, s) where u starts with $u_0, \ldots, u_j$ and $\sum_{i=0}^{k} s_i = t'$. We have that

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

□

Appendix B. A Comparison of the Existing Definitions of CTMDPs

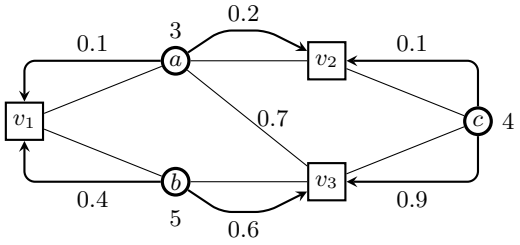
As we already mentioned in Section 2, our definition of CTG (and hence also CTMDP) is somewhat different from the definition of CTMDP used in [**?** **?**]. To prevent misunderstandings, we discuss the issue in greater detail in here and show that the two formalisms are in fact equivalent. First, let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

Definition B.30. A CTMDP is a triple $\mathcal{M} = (S, A, \mathbf{R})$, where S 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*.

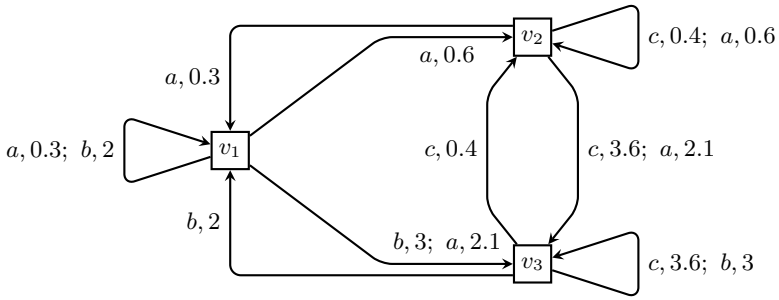
A CTMDP $\mathcal{M} = (S, A, \mathbf{R})$ can be depicted as a graph where S is the set of vertices and $s \to s'$ is an edge labeled by (a, r) iff $\mathbf{R}(s, a, s') = r > 0$. The conditional probability of selecting the edge $s \xrightarrow{(a,r)} s'$, under the condition that the action a is used, 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 the set of vertices is S, 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.30) looks as follows:



However, there is one subtle issue regarding strategies. In [**?** **?**], a strategy (controller) selects an action in every vertex. The selection may depend on the history of a play. In [**?** **?**], 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.

Appendix C. Technical Proofs

Appendix C.1. Proofs of Claim 3.7 and Claim 3.8

Claim 3.7. *$\mathcal{A}$ is a fixed point of $\mathcal{V}$.*

Proof. 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 τ^a a strategy defined by $\tau^a(wu) := \tau(vawu)$. Note that $\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} P^{\sigma,\pi}(\cdot,\cdot) = \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} P^{\sigma^a,\pi^a}(\cdot,\cdot)$ for any $a \in A$.

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}(\mathbf{R}(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}(\mathbf{R}(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}(\mathbf{R}(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}(\mathbf{R}(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}(v, \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}(\mathbf{R}(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} \pi(v)(a) \sum_{u \in V} \mathbf{P}(a)(u) \sum_{\substack{\mathbf{j} \in \mathcal{H} \\ \mathbf{j}(\mathbf{R}(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}(\mathbf{R}(a))>0}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot \pi(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}(\mathbf{R}(a))>0}} \pi(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}(v, \mathbf{j})
\end{aligned}$$

□

Claim 3.8. *For every $\sigma \in \Sigma$, $k \geq 0$, $\mathbf{i} \in \mathcal{H}$, $v \in V$, $\varepsilon \geq 0$, we have*

$$\mathcal{R}_k^\sigma(\mathbf{i}, v) \leq \mu\mathcal{V}(\mathbf{i}, v) + \sum_{j=1}^{k} \frac{\varepsilon}{2^{|\mathbf{i}|+j}}$$

Proof. For $v \in T$ we have

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

Assume that $v \notin T$. We proceed by induction on k. For $k = 0$ we have

$$\mathcal{R}_k^\sigma(\mathbf{i}, v) = 0 \leq \mu\mathcal{V}(\mathbf{i}, v)$$

For the induction step, first assume that $v \in V_\Box \setminus T$

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

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

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

Appendix C.2. Proof of Lemma 3.12

Lemma 3.12. *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*

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

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 ε-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)

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

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

Claim C.34. 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

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

by assumptions. □

ad 2. We proceed similarly as in the proof of Theorem 3.6 (we also use some notation of the proof of Theorem 3.6). 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 $\alpha \in Run_{G(u,\sigma,\pi)}$ 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 σ and π and assuming that the history matches $\mathbf{i}$.

To define the CD strategies σ^k and π^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 σ^k and π^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 τ 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 claim.

Claim C.35. Let $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$. Then

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

In particular, the strategies σ^k and π^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 (C.1). Using the notation of the proof of Theorem 3.6, given a strategy $\tau \in \Sigma \cup \Pi$ and $a \in A$, we denote by τ^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 (C.2) can be proved in a similar manner.

The claim follows from the following

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

□

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

Appendix C.3. Proof of Claim 4.19

Claim 4.19. *A strategy is always greedy on s steps iff it uses transitions from $\mathbf{E}_s$ only (as defined by Algorithm 1).*

Proof. For the 'if' direction, we prove by induction on n that

$$\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\big(\overrightarrow{\mathcal{P}}_v^{\sigma,\pi}\big)_{(1,\dots,n)} = \min_{\pi\in\Pi}\big(\overrightarrow{\mathcal{P}}_v^{\bar\sigma,\pi}\big)_{(1,\dots,n)} = \max_{\sigma\in\Sigma}\big(\overrightarrow{\mathcal{P}}_v^{\sigma,\bar\pi}\big)_{(1,\dots,n)}$$

for all strategies $\bar\sigma \in \Sigma$ and $\bar\pi \in \Pi$ that use edges from $\mathbf{E}_n$ only. It is sufficient to prove that

$$\big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_v^{\sigma,\pi}\big)_n \overset{(1)}{=} R_n(v) \overset{(2)}{=} \min_{\pi\in\Pi}\big(\overrightarrow{\mathcal{P}}_v^{\bar\sigma,\pi}\big)_n$$

for every strategy $\bar\sigma \in \Sigma$ that uses edges from $\mathbf{E}_n$ only. (The minimizing part is dual.) The case $n = 0$ is trivial. Now consider $n+1$. For $v \in V_\Box \setminus T$,

$$\begin{aligned}
\big(\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_v^{\bar\sigma,\pi}\big)_{n+1} &= \sum_{u\in V}\bar\sigma(0,v)(u)\big(\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_u^{\bar\sigma[1],\pi}\big)_n \\
\text{by IH (2) and } \mathbf{E}_{n+1}\subseteq\mathbf{E}_n \quad &= \sum_{u\in V}\bar\sigma(0,v)(u)\cdot R_n(u) \\
\bar\sigma \text{ uses } \mathbf{E}_{n+1} \text{ only} \quad &= \max_{a\in\mathbf{E}_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot R_n(u) \quad = \quad (*) \\
\text{by IH (1)} \quad &= \max_{a\in\mathbf{E}_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot\big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_u^{\sigma,\pi}\big)_n \\
\text{by IH} \quad &= \max_{a\in A_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot\big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_u^{\sigma,\pi}\big)_n \\
&= \big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_v^{\sigma,\pi}\big)_{n+1}
\end{aligned}$$

where $A_n(v)$ is the set of all edges going from v that any strategy always greedy on n steps can choose. I.e. it is the desired abstractly defined set of greedy edges, which is equal to the computed set $\mathbf{E}_n(v)$ by the induction hypothesis. Since $(*) = R_{n+1}(v)$, the equality with the first and the last expression proves the claim. Similarly for $v \in V_\Diamond \setminus T$,

$$\begin{aligned}
\big(\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_v^{\bar\sigma,\pi}\big)_{n+1} &= \min_{a\in A_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot\big(\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_u^{\bar\sigma[1],\pi}\big)_n \\
\text{by IH (2) and } \mathbf{E}_{n+1}\subseteq\mathbf{E}_n \quad &= \min_{a\in A_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot R_n(u) \\
\text{by IH for the minimizing part} \quad &= \min_{a\in\mathbf{E}_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot R_n(u) \quad = \quad (**) \\
\text{by IH (1)} \quad &= \min_{a\in\mathbf{E}(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot\big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_u^{\sigma,\pi}\big)_n \\
\text{by IH for the minimizing part} \quad &= \max_{a\in A_n(v)}\sum_{u\in V}\mathbf{P}(a)(u)\cdot\big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_u^{\sigma,\pi}\big)_n \\
&= \big(\max_{\sigma\in\Sigma}\min_{\pi\in\Pi}\overrightarrow{\mathcal{P}}_v^{\sigma,\pi}\big)_{n+1}
\end{aligned}$$

where $(**) = R_{n+1}(v)$. The case with $v \in T$ is trivial as states in T are absorbing.

We prove the "only if" direction by contraposition. If a strategy τ uses a transition $a \in \mathbf{E} \setminus \mathbf{E}_s$ in v then there is $i \leq s$ such that a has been cut off in the ith step. Therefore a did not realize the i steps optimum (equal to $R_i(v)$). Hence τ is not greedy on n steps. □

Appendix C.4. Proof of Lemma 5.24

Lemma 5.24. *The strategies σ_ε and π_ε are ε-optimal.*

Proof. We use some notation of the proof of Theorem 3.6. 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 $\alpha \in Run_{G(u,\sigma,\pi)}$ 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 $(\sigma, \pi) \in \Sigma \times \Pi$, $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, 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 σ and π and assuming that the history already matches $\mathbf{i}$. We have shown in the proof of Claim C.35 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 $\bar{P}^{\sigma,\pi}(\mathbf{0}, v) = \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}_{\leq k}(T))$ and thus $\bar{R}(\mathbf{0}, v) = val^k(v)$, 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 σ_ε and π_ε 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 $\left(\frac{\varepsilon}{2}\right)^{2|\mathbf{i}|+1}$ (Claim C.38 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 $\left(\frac{\varepsilon}{2}\right)^{2|\mathbf{i}|+1}$ as well (Claim C.39 below). In other words, σ_ε and π_ε are $\frac{\varepsilon}{2}$-optimal strategies in the k-step game. Together with the assumptions imposed on k we obtain that σ_ε and π_ε are ε-optimal strategies.

For $n \geq 0$, we denote by err_n the number $\left(\frac{\varepsilon}{2}\right)^{2n+1}$.

Claim C.38. For all $\mathbf{i} \in \mathcal{H}$ and $v \in V$ we have

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

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

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

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.

□

We denote by Σ_{CD} and Π_{CD} the sets of all CD strategies of Σ and Π, respectively. Recall that 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 τ 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

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

Claim C.39. Let $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$. We have

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

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

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

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

This proves that σ_ε and π_ε are ε-optimal, since the absolute error is smaller than the relative error as the probabilities are at most 1. □