Quantitative model-checking of controlled discrete-time Markov processes[☆]

Ilya Tkachev[a], Alexandru Mereacre[b], Joost-Pieter Katoen[c], Alessandro Abate[b,*]

[a]*Delft Center for Systems & Control, Delft University of Technology, The Netherlands*
[b]*Department of Computer Science, University of Oxford, United Kingdom*
[c]*Software Modeling and Verification Group, RWTH Aachen University, Germany*

Abstract

This paper focuses on optimizing probabilities of events of interest defined over general controlled discrete-time Markov processes. It is shown that the optimization over a wide class of ω-regular properties can be reduced to the solution of one of two fundamental problems: reachability and repeated reachability. We provide a comprehensive study of the former problem and an initial characterisation of the (much more involved) latter problem. A case study elucidates concepts and techniques.

1. Introduction

Stochastic control models have been widely investigated and employed in numerous applications in different areas such as finance, biology, power networks, etc. – see [HLL96, Chapter 1] or [Mey08] for examples. Under discrete time semantics, a natural way to model probabilistic behavior allowing for the presence of control inputs is to employ the framework of controlled discrete-time Markov processes (`cdt`-MP), also known as general Markov Decision Processes (MDP) [FS02]. In this modeling formalism, given the current state of a system and the control action provided by an external agent, the distribution of the next state is uniquely (deterministically) determined, which also entails the Markovian structure of the model. In turn, the choice of the control action itself may depend on the complete history of state and control observations, and can be randomized. The decision rule of the agent, which assigns to the history observation a choice of the next action, is called the policy. Unlike known results over finite-state models [BK08], in this work we deal with general models evolving over uncountable spaces.

[☆]This work has been supported by the European Commission STREP project MoVeS 257005. The authors are grateful to Bill Sudderth for helpful discussions on gambling theory.

[*]Corresponding author

Email addresses: `i.tkachev@tudelft.nl` (Ilya Tkachev), `mereacre@gmail.com` (Alexandru Mereacre), `katoen@cs.rwth-aachen.de` (Joost-Pieter Katoen), `alessandro.abate@cs.ox.ac.uk` (Alessandro Abate)

A generic optimization problem over a `cdt`-MP is the following: given a performance criterion whose value is uniquely determined by a chosen policy [Fei83], optimize (maximize or minimize) the value of this criterion over the given class of policies, and determine (if possible) the policy (or set of equivalent policies) corresponding to the optimal value. In the literature a wide range of performance criteria have been studied – see e.g. [ABFG$^+$93, Section 3] for remarks on the historical development of the topic – among them the discounted cost (DC), the total cost (TC), and the average cost (AC). All these criteria present an additive structure, which allows for the solution by means of dynamic programming (DP) [Bel54], namely a backward-recursive procedure that computes the optimal control action by balancing the present value of the cost and the expected future cost caused by the choice of such an action. The DP approach has led to a rich theory for such criteria – see [BS78] for an overview on the DC and TC, and [ABFG$^+$93] for a survey on AC. Unfortunately similar results for other sorts of criteria are much less comprehensive, the focus in the literature being more on qualitative analysis, e.g. determining which policy classes are sufficient to focus on, and no general solution techniques have been developed to the best of our knowledge. This in particular is the case when one wants to optimize the probability of a given event, examples of the latter being "the state trajectory never leaves the safe set S" or "the state trajectory eventually reaches the goal set G without leaving the safe set S beforehand". Instances of these problems have been studied in isolation [MPS91, MS96b], however no comprehensive treatment for this general class of problems has been given.

In this work we apply methods grounded on modal logic and on automata theory for the following two purposes: first, we develop a framework to quantitatively define a class of performance criteria of interest, encompassing the instances discussed above; second, we solve optimization problems over such criteria in a unified way. More specifically, we propose to express events as formulae within a linear temporal logic (LTL) [BK08, Chapter 5], encompassing intuitive specifications on the model that are related to sentences in natural languages. We further show that such formulae can be recast as automata: simple dynamical systems endowed with a logical structure given by their acceptance conditions [BK08, Chapter 4]. We prove that the optimization of any given event expressed as an automaton over the original `cdt`-MP model can be reduced to one of two fundamental problems, namely reachability or repeated reachability: the former requires visiting a goal set at least once, whereas the latter requires infinitely many visits to the goal set.

The reachability problem over `cdt`-MP has been recently studied e.g. in [APLS08, SL10, CCL11], however the results have either required restrictive conditions the on model or focused on special cases of the problem, for instance when only Markov (history-independent) policies are allowed. In contrast, here we consider the most general setting for the reachability problem, and we provide a complete treatment of the problem under conditions on the model being as mild as possible: this is considered to be the core of our contribution. For example, up to our knowledge we are the first to give a comprehensive study of the unbounded-horizon reachability over `cdt`-MP, providing Lyapunov-like techniques for its solution. In order to obtain these results, we show that the reachability performance criterion can be expressed as a TC one over a modified `cdt`-MP, which allows us extending the rich

theory for the latter criterion to the reachability case. Unfortunately, we are not able to give a comparable study of the repeated reachability problem, however we extend results from gambling theory [MS96b] to characterize the DP formulation for this problem, and propose a solution using Lyapunov-like excessive functions in the special case when the system possesses certain stability properties. Gambling theory is an area interested in optimising probabilistic decision models that closely remind MDPs, however with the main goal of optimizing gambler's behaviour. For this goal, the performance criteria studied in gambling theory (final value of the capital) are a bit different from classical problems for MDPs (discounted additive payoff), which we exploit when focusing on the repeated reachability problem.

An approach to the optimal control of `cdt`-MP based on LTL and automata has been developed for finite-state and finite-action models in the model-checking literature [BK08, Section 10.6]: however that setup does not deal with technical issues (e.g. measurability) that appear in our context. Due to this reason, our contribution can be considered from two perspectives. For readers familiar with formal methods in control [Tab09] and model-checking [BK08], we extend the model-checking techniques from finite `cdt`-MP to a general class of models, whereas for readers experienced in classical stochastic optimal control we propose a novel formulation and solution of the problem of optimization of probabilities of classes of events of interest.

The rest of the paper is organized as follows. The model description and the problem formulation are given in Section 2, which also puts forward the result on reduction of the general problem to either of two fundamental ones: reachability or repeated reachability. Section 3 is devoted to the former case, whereas Section 4 is focused on the latter instance. We give an elucidating numerical case study in Section 5, and the paper is concluded in Section 6. The notation and the background in analysis and measure theory are provided in the Appendix.

2. Models and problem formulation

2.1. Model syntax and semantics

The models considered in this work are known as controlled discrete-time Markov processes (`cdt`-MP), formalised next. A `cdt`-MP is a discrete-time stochastic model with a specific transition structure: the distribution of the next state of the process is completely determined by the current state and the current choice of the control action. These models are alternatively known in the literature as controlled Markov models [HLL96], general Markov Decision Processes (MDP) [Put94] or gambling houses [MS96a]. There are often slight variations in the their definition: the one we give here is inspired by the Borel model introduced in [BS78, Chapters 8, 9]. Details on notation can be found in the Appendix.

Definition 1 (Controlled discrete-time Markov process)**.** *A* `cdt`-MP *is a tuple* $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$*, where* X *and* U *are non-empty Borel spaces,* $\mathbb{K}$ *is an analytic subset of* $X \times U$*, and* $\mathsf{T} \in \mathscr{B}(X|X \times U)$ *is a stochastic kernel.*

The `cdt`*-MP* $\mathfrak{D}$ *is called* continuous *if* U *is a compact Borel space,* $\mathbb{K}$ *is a closed subset of* $X \times U$ *and the restriction* $\mathsf{T}|_{\mathbb{K}}$ *is a continuous kernel.*

Given a cdt-MP $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ we say that X is its state space, U is the action space, $\mathbb{K}_x$ are the actions that are feasible at state $x \in X$, and T is the transition kernel. The latter induces several operators that act on functions defined over the state space. For any $\mu \in \mathscr{U}(U|X)$ and any function $f \in \mathrm{b}\mathscr{U}(X)$ we define

$$\mathsf{T}^{\mu} f(x) := \int_{X \times U} f(x') \mathsf{T}(\mathrm{d}x'|x,u) \mu(\mathrm{d}u|x).$$

In particular, when $\mu = \delta_u$ is a constant kernel, where u is some element of U, we simply write T^u rather than T^{δ_u}. Clearly, it holds that $\mathsf{T}^u 1_A(x) = \mathsf{T}(A|x,u)$ for all $x \in X$, $u \in U$ and any $A \in \mathscr{B}(X)$. Furthermore, [BS78, Proposition 7.46] implies that T^{μ} maps $\mathrm{b}\mathscr{U}(X)$ to itself. We also consider the following operators:

$$\mathsf{T}^* f(x) := \sup_{u \in U} \mathsf{T}^u f(x), \qquad \mathsf{T}_* f(x) := \inf_{u \in U} \mathsf{T}^u f(x).$$

If $f \in \mathrm{b}\mathscr{A}_*(X)$, then $\mathsf{T}^u f \in \mathrm{b}\mathscr{A}_*(X)$ thanks to [BS78, Proposition 7.48]. Furthermore, it follows from [BS78, Proposition 7.47] that $\mathsf{T}_* f(x) \in \mathrm{b}\mathscr{A}_*(X)$ as well and, as a result, the operator T_* maps the space $\mathrm{b}\mathscr{A}_*(X)$ into itself. Similar arguments show that the operator T^* maps the space $\mathrm{b}\mathscr{A}^*(X)$ into itself.

The semantics of the cdt-MP $\mathfrak{D}$ is given as follows: at any time instant $k \in \mathbb{N}_0$, if the state of $\mathfrak{D}$ is $x_k \in X$ and the action $u_k \in K_{x_k}$ is chosen, then the new state x_{k+1} is a random variable distributed according to the following law:

$$x_{k+1} \sim \mathsf{T}(\cdot|x_k, u_k). \tag{2.1}$$

As a known example, every stochastic difference equation of the form

$$x_{k+1} = F(x_k, u_k, \xi_k), \tag{2.2}$$

where $(\xi_k)_{k \in \mathbb{N}_0}$ is a sequence of iid random variables and the map $F : X \times U \times \mathbb{R} \to X$ is Borel measurable, can be represented as in (2.1). In this case the kernel T can be expressed via the map F as

$$\mathsf{T}(B|x,u) = \nu(\{\xi \in \mathbb{R} : F(x,u,\xi) \in B\}),$$

for any $B \in \mathscr{B}(X)$, where ν is the distribution of ξ_0. On the other hand, the converse statement also holds true, though there is in general no constructive method to derive an F from a given T [HLL96, Section 2.3]. Although (2.2) may be more intuitive or familiar, the representation of the dynamics as in (2.1) is preferred in this work. Note also that if F as in (2.2) is such that $F(\cdot, \xi) : X \times U \to X$ is a continuous map, then the corresponding kernel is continuous as well [HLL96, Example C.7].

A formal definition of the evolution of a cdt-MP is given by its paths and by the corresponding probability measures on the path space. More precisely:

Definition 2 (Semantics of cdt-MP)**.** *Given a* cdt*-MP* $\mathfrak{D}$*, its* infinite path *is an infinite sequence*

$$h = (x_0, u_0, x_1, u_1, \dots), \tag{2.3}$$

where $x_k \in X$ are the state coordinates and $u_k \in U$ are the action coordinates of the path, $k \in \mathbb{N}_0$. The space of all infinite paths is denoted by $H := (X \times U)^{\mathbb{N}_0}$ (cf. notations in Appendix) and is called the canonical sample space *of the* cdt*-MP $\mathfrak{D}$.*

For $n \in \mathbb{N}_0$, a finite n*-path h_n is a finite prefix of an infinite path ending in a state:*

$$h_n = (x_0, u_0, \dots, x_{n-1}, u_{n-1}, x_n), \tag{2.4}$$

where $x_k \in X$ and $u_k \in U$. The space of all n-paths is denoted by $H_n = (X \times U)^n \times X$.

Infinite paths of cdt-MP are mostly used to introduce certain performance criteria over the model, whereas finite n-paths naturally serve as the history of observation (the past information) available up to instant n. Due to this reason, we use notation H and H_n for the spaces of paths, and below we often refer to finite paths as *histories*.

Similarly to [ABFG+93], we define the state, action, and information processes on a sample space H. They are denoted respectively by $(\mathbf{x}_n)_{n\in\mathbb{N}_0}$, $(\mathbf{u}_n)_{n\in\mathbb{N}_0}$ and $(\mathbf{h}_n)_{n\in\mathbb{N}_0}$, and are defined by the following projections on spaces X, U and H_n:

$$\mathbf{x}_n(h) := x_n, \quad \mathbf{u}_n(h) := u_n, \quad \mathbf{h}_n(h) := (x_0, u_0, \dots, x_{n-1}, u_{n-1}, x_n), \quad n \in \mathbb{N}_0,$$

for any $h \in H$ as per (2.3). Notice that it may happen that $\mathbf{u}_k(h) \notin \mathbb{K}_{\mathbf{x}_k(h)}$, which reflects action coordinates that are not feasible: this is allowed for technical reasons and later we show that the corresponding paths are of measure zero.

When dealing with stochastic processes, questions of measurability are crucial to render objects well-defined. This in particular applies to the choice of action u_n at time n, given the history h_n, and is formalized using the notion of policy.[1]

Definition 3 (Policy for cdt-MP)**.** *Given a* cdt*-MP $\mathfrak{D}$, a* policy *is a sequence $\pi = (\pi_n)_{n\in\mathbb{N}_0}$ of universally measurable kernels $\pi_n \in \mathscr{U}(U|H_n)$, which is such that for any h_n as in (2.4) it holds that*

$$\pi_n(K_{x_n}|h_n) = 1. \tag{2.5}$$

The class of all policies of $\mathfrak{D}$ is denoted by Π.

Notice that (2.5) implies that policies are only allowed to select actions among the currently feasible ones. Once a policy $\pi \in \Pi$ and an initial distribution $\alpha \in \mathscr{P}(X)$ are fixed, the behavior of a cdt-MP $\mathfrak{D}$ is completely characterized by the probability measure $\mathsf{P}_\alpha^\pi \in \mathscr{P}(H)$ over the path space H. This measure is uniquely defined by

$$\begin{aligned}\int_H f \,\mathrm{d}\mathsf{P}_\alpha^\pi = \int_X \int_U \int_X \dots \int_X \int_U f(x_0, u_0, x_1, \dots, x_{n-1}, u_{n-1}, x_n, \dots) \\ \times \mathsf{T}(\mathrm{d}x_n|x_{n-1}, u_{n-1})\pi_{n-1}(\mathrm{d}u_{n-1}|x_0, u_0, \dots, x_{n-1}) \\ \times \mathsf{T}(\mathrm{d}x_{n-1}|x_{n-2}, u_{n-2}) \cdots \mathsf{T}(\mathrm{d}x_1|x_0, u_0)\pi_0(\mathrm{d}u_0|x_0)\alpha(\mathrm{d}x_0),\end{aligned} \tag{2.6}$$

[1] Policies are also known as strategies or feedback controls [MS96a], or alternatively as schedulers or adversaries [BK08]. Whilst the earlier ones are usually employed to optimise given criteria, in the latter case they are used to resolve non-determinism in non-deterministic stochastic models, such as probabilistic automata [SL95].

for any bounded $\mathbf{h}_n$-measurable function $f : H \to \mathbb{R}$ [BS78, Sections 8.1, 9.1]. In particular, $\mathsf{P}_\alpha^\pi(\mathbb{K}^{\mathbb{N}_0}) = 1$ so that (as we anticipated) the probability of paths containing non-feasible actions is equal to zero. Moreover, by taking f to be an appropriate indicator function, for any sets $B \in \mathscr{B}(X)$ and $C \in \mathscr{B}(U)$ we obtain the following equalities that hold P_α^π-a.s.:

$$\mathsf{P}_\alpha^\pi(\mathbf{x}_0 \in B) = \alpha(B), \tag{2.7}$$

$$\mathsf{P}_\alpha^\pi(\mathbf{u}_n \in C | \mathbf{h}_n) = \pi_n(C|\mathbf{h}_n), \tag{2.8}$$

$$\mathsf{P}_\alpha^\pi(\mathbf{x}_{n+1} \in B | \mathbf{h}_n, \mathbf{u}_n) = \mathsf{T}(B|\mathbf{x}_n, \mathbf{u}_n). \tag{2.9}$$

As a result, the probability measure P_α^π captures all the intuitive features about the behavior of the `cdt`-MP $\mathfrak{D}$ under the selected policy π and given the initial distribution α. In particular, (2.9) implies that the distribution of $\mathbf{x}_{n+1}$ only depends on $\mathbf{x}_n$ and $\mathbf{u}_n$, rather than on the whole history $\mathbf{h}_n$. Note, however, that the chosen control action $\mathbf{u}_n$ itself can depend on the history rather than only on the current state $\mathbf{x}_n$. We say that $\left(H, \mathscr{B}(H), (\mathsf{P}_\alpha^\pi)_{\alpha \in \mathscr{P}(X)}^{\pi \in \Pi}\right)$ is the canonical probability space for the `cdt`-MP $\mathfrak{D}$.[2] Finally, as a shorthand, the notation P_x^π is used in place of $\mathsf{P}_{\delta_x}^\pi$.

We conclude the discussion by highlighting the following important classes of policies over the `cdt`-MP $\mathfrak{D}$.

- Π_M – the class of *Markov* policies. A policy $\pi \in \Pi$ is called Markov if for any $n \in \mathbb{N}_0$ the measure $\pi_n(h_n)$ depends only on x_n for any finite path $h_n \in H_n$ as per (2.4). More precisely: for all $n \in \mathbb{N}_0$ and $x_n \in X$

  $$\pi_n(h', x_n) = \pi_n(h'', x_n), \quad \forall h', h'' \in H_{n-1} \times U.$$

  This means that a Markov policy selects an action solely based on the information about the current state, rather than on the whole available history.

- $\Pi_S \subseteq \Pi_M$ – the class of *stationary* policies. A Markov policy π is called stationary if $\pi_n(x) = \pi_{n+1}(x)$ for all $n \in \mathbb{N}_0$ and $x \in X$. Thus, stationary policies are time-independent.

- Π^D – the class of *deterministic* policies. A policy $\pi \in \Pi$ is called deterministic if $\pi_n = \delta_f$ for some universally measurable map $f : H_n \to U$.

- $\Pi_S^D \subseteq \Pi_M^D \subseteq \Pi^D$ – classes of deterministic Markov and deterministic stationary policies. Such classes are defined by $\Pi_M^D = \Pi^D \cap \Pi_M$ and $\Pi_S^D = \Pi^D \cap \Pi_S$.

We refer to any map $f : X \to U$ satisfying $\mathsf{Gr}[f] \subseteq \mathbb{K}$ as a selector (from $\mathbb{K}$), whereas a stochastic kernel $\mu \in \mathscr{U}(U|X)$ satisfying $\mu(\mathbb{K}_x|x) = 1$ for all $x \in X$ is called a randomized selector. Clearly, the existence of the former (of the latter) is equivalent to the statement that the policy class Π_S^D (Π_S) is not empty. Notice that Π_S^D is the smallest among the classes of policies introduced above, and since $\mathbb{K}$ is

[2] We slightly abuse the notation here since in fact this is a family of probability spaces parameterized by $\alpha \in \mathscr{P}(X)$ and $\pi \in \Pi$, rather than a single probability space.

analytic, it admits an analytically measurable selector, namely it contains the graph of an analytically measurable map $\mathsf{k} : X \to U$ [BS78, Proposition 7.49]. As a result, Π_S^D is not empty, hence neither are all other classes.

2.2. Example: a small power network

In order to elucidate the concepts introduced above, let us discuss the following example, modified from the one in [TMKA13]. Figure 2.2 schematically depicts the setup.

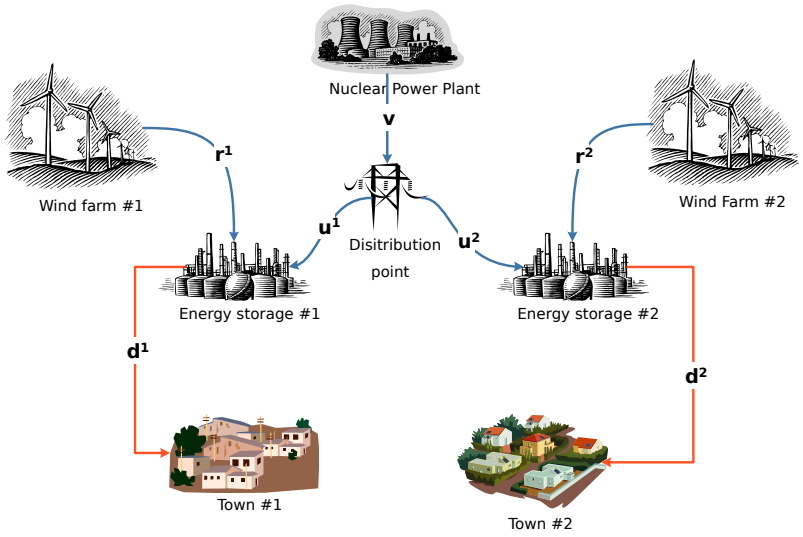


Figure 1: Case study: a power network consisting of two sub-networks.

Consider a simple, abstract power network consisting of two aggregated consumers (e.g. small towns), each of which benefits from a separate generator of renewable energy (e.g. a wind farm) and a separate energy storage. Suppose that in addition there is a shared polluting power generator, such as a nuclear power plant. The energy flow is assumed to be stochastic, in particular due to the production deriving from the wind farms. The energy output of the nuclear power plant is less volatile and larger. Within this setup one requires that the energy supply is greater than the energy demand, or imposes additional requirements on the energy levels. The available control is the total load on the nuclear power plant, as well as its distribution over the two consumers. More precisely, the model is given as follows:

$$\mathbf{x}_{k+1}^i = c \cdot \left(\mathbf{x}_k^i + \mathbf{u}_k^i \cdot \mathbf{v}_k \cdot p_k + r_k^i - d_k^i\right) \wedge M \vee 0, \tag{2.10}$$

where $\mathbf{x}_k^i \in [0, M]$ is the energy level in the subnetwork $i \in \{1, 2\}$ at the discrete time instance $k \in \mathbb{N}_0$ (encompassing daily updates), and $M > 0$ is the maximal

storage capacity. The constant $c \in (0,1]$ is the reserve rate of the stored energy, $\mathbf{v}_k \in [v_{\min}, 1]$ is the load on the nuclear power plant and $\mathbf{u}_k^i \in [0,1]$ is the share of energy produced by the nuclear power plant that is supplied to the subnetwork i, so that $\mathbf{u}_k^1 + \mathbf{u}_k^2 \equiv 1$. As we assume that it is not possible to switch the nuclear power plant off, $v_{\min}$ is the minimal load on the plant. The noise is represented by a sequence of iid random variables accounting for uncertainty on the nuclear power plant actual production p_k, the wind farm production $r_k = (r_k^1, r_k^2)$, and the total local demand $d_k = (d_k^1, d_k^2)$. Note that r_k^1 and r_k^2 are not necessarily independent (coupling can be due to weather), and neither are the demand variables for the subnetworks d_k^1 and d_k^2.

A `cdt`-MP model for the dynamics above is given considering a state space $X = [0, M]^2$ with $\mathbf{x}_k = (\mathbf{x}_k^1, \mathbf{x}_k^2)$, control space $U = [0,1]^2$ with $\mathbf{u}_k = (\mathbf{v}_k, \mathbf{u}_k^1)$, and control actions that are always feasible (namely $\mathbb{K} = X \times U$), and a transition kernel induced by the stochastic difference equation (2.10). Goals for control synthesis are discussed shortly, whereas the analysis of the model and the synthesis problem are presented in Section 5.

2.3. Problem formulation

The framework of `cdt`-MP is often used in an optimization context. In particular, one of the most prominent questions to answer is the following: what is the maximal achievable value of a given performance measure, and can a control policy that achieves such a value be derived? Clearly, the answer crucially depends on the chosen optimisation goal: since this choice is quite broad in the literature on `cdt`-MP, let us discuss some important cases.[3]

We do not consider multi-objective optimization where the performance criterion has a partial order on its co-domain (see e.g. [Bor91]), and instead focus on numerical criteria, namely measures taking values on $\mathbb{R}$. Arguably one of the most general approaches to the definition of numerical performance criteria over `cdt`-MP has been considered in [Fei83]. There, the focus is on the space of *strategic* measures given by $\mathsf{S} := \{\mathsf{P}_\alpha^\pi | \pi \in \Pi, \alpha \in \mathscr{P}(X)\}$, and the criterion is simply any function $f : \mathsf{S} \to \mathbb{R}$. A slightly more specific class of criteria is related to the concept of the *expected utility* [Kre77a, Kre77b, Kre78]. A utility is any history-dependent random variable $J : H \to \mathbb{R}$ and the corresponding performance is defined to be its expected value $\mathsf{M}^\pi(\alpha; J) := \mathsf{P}_\alpha^\pi[J]$, where P_α^π is a probability measure over paths that depends on a policy and an initial distribution. Clearly, the expected utility criterion is a special case of the former, since to any utility J one can assign a function $f_J : \mathsf{S} \to \bar{\mathbb{R}}$ by defining $f_J(p) := p[J]$ for any $p \in \mathsf{S}$. Research on these criteria has led to strong theoretical results, e.g. on the characterization of classes of optimal policies. On the other hand, the generality of the problems did not allow for specific results related to the computability of the optimal solutions. Due to this reason, more specific performance criteria have attracted a significant interest, in particular the *discounted cost* (DC) and the *average cost* (AC) criteria [ABFG+93].

[3] A comprehensive survey on different performance criteria, as well as on the general development of the theory of `cdt`-MP, is given in [ABFG+93, Section 3].

Consider some universally measurable *cost function* $c : \mathbb{K} \to \bar{\mathbb{R}}$ and define

$$\mathsf{DC}^{\pi}_{n,\gamma}(x) := \mathsf{P}^{\pi}_{x}\left[\sum_{k=0}^{n} \gamma^{k} c(\mathbf{x}_k, \mathbf{u}_k)\right],$$

$$\mathsf{AC}^{\pi}(x) := \limsup_{n\to\infty} \frac{1}{n} \mathsf{P}^{\pi}_{x}\left[\sum_{k=0}^{n} c(\mathbf{x}_k, \mathbf{u}_k)\right]$$

where $\gamma \in (0,1]$ is the *discounting factor* and $n \in \bar{\mathbb{N}}_0$ is the *time horizon*. The DC is clearly a special case of the expected utility criterion. In general it is not possible to express AC as an expected utility, but clearly it is still a function of strategic measures and thus belongs to the class of criteria considered in [Fei83]. Furthermore, with focus on the DC, the case $\gamma = 1$ is often referred to as the *total cost* (TC) criterion or, alternatively, the *additive cost*. These problems are extensively studied in the literature: see e.g. [BS78, HLL96] for the DC, and [ABFG+93] for the AC.

The focus of this paper is on the probabilities associated to certain events defined over the paths of the cdt-MP. More precisely, let $A \in \mathscr{B}(H)$ be some set of desired path behaviors of the cdt-MP, and consider a performance criterion to be $\mathsf{P}^{\pi}_{\alpha}(A)$. Clearly, this is still a special case of the expected utility criterion, with the utility given by 1_A, and thus general results apply. However, if we focus on a certain class of events, rather than considering all possible elements of $\mathscr{B}(H)$, it is possible to obtain much stronger results in terms of characterization and of computability. More specifically, we exploit the known approach in formal methods [BK08] to treat any event as a property (or a *specification*) over paths of a cdt-MP. Such a property is further expressed as a simple dynamical model satisfying it. This technique has been widely employed to study cdt-MP models over finite state and action spaces [CY98], leading to analytical solutions for that setup. However, the developed methods appear to be crucially dependent on the discrete structure of finite cdt-MP and thus are not fully applicable to the general case. The aim of this work is to develop new techniques to tackle this problem over *general* cdt-MP.

Before we describe the class of events of interest, let us introduce some notation for the expected utility criterion. Given an initial distribution $\alpha \in \mathscr{P}(X)$, a policy $\pi \in \Pi$ and a random variable $f \in \mathrm{b}\mathscr{B}(H)$, we denote $\mathsf{M}^{\pi}(\alpha; f) := \mathsf{P}^{\pi}_{\alpha}[f]$. In the particular case when $f = 1_A$ for some $A \in \mathscr{B}(H)$ or $\alpha = \delta_x$, we simply write $\mathsf{M}^{\pi}(\alpha; A)$ or $\mathsf{M}^{\pi}(x; A)$. The optimal expected utility functions are defined as

$$\mathsf{M}^{*}(\alpha; f) := \sup_{\pi\in\Pi} \mathsf{M}^{\pi}(\alpha; f), \qquad \mathsf{M}_{*}(\alpha; f) := \inf_{\pi\in\Pi} \mathsf{M}^{\pi}(\alpha; f).$$

In order to formulate the problem, we need to specify the class of events we focus on. Recall the power network model from Section 2.2, and consider the following tasks:

- keep the energy levels always within specified target levels;
- test the network as follows: reach an energy level above the target value over the first subnetwork, and while keeping it there, reach the same energy level

over the second network, or attain dual goal. In addition, avoid blackouts, that is never allow an energy level of either of the subnetworks to reach the zero level.

The first task corresponds to a *safety* problem, which can be easily characterized using canonical probabilistic tools or the concept of the first hitting time. On the other hand, the second task is more complicated, even in its qualitative description. For this purpose we introduce a modal logic called Linear Temporal Logic (LTL), which is useful in the following two aspects. First of all, it provides *"a very intuitive but mathematically precise notation"* [BK08, Section 5.1] to deal with a large class of complex and interesting events. Secondly, LTL allows reducing the optimization problems for any of such events to one of the following two fundamental problems: *reachability*, requiring visiting a specified target set at least once; or *constrained repeated reachability*, requiring visiting a target set infinitely often and visiting an unsafe set only finitely often.

LTL is introduced using its *grammar*, namely the set of rules determining the construction of LTL formulae. The meaning of each formula (that is, the event corresponding to the formula) is formalized by the LTL *semantics*. It is canonical to introduce the latter not directly over the state space, but rather using the concept of *labels*, namely discrete observations of states that range over some finite set called the *alphabet*. Alternatively, one can think of assigning some distinguishable sets to the state space. Intuitively, when one says that $x \in A$ it may be considered as an implicit assignment of an abstract label "A" to point x.

Consider a finite set of atomic propositions and its power set Σ, which is referred to as the alphabet. Elements of Σ are called *letters*, whereas finite or infinite sequences of letters are called *words*. Let us denote by Σ^ω the space of infinite words; by infinite *language* over Σ we mean any collection of infinite words over the alphabet Σ. All the languages we consider in this paper are assumed to be infinite, i.e. we say that ϕ is a language to mean $\phi \subseteq \Sigma^\omega$. If $\phi \in \mathscr{B}(\Sigma^\omega)$ we say that ϕ is a measurable language. In particular, it follows from [Var85, Proposition 2.3] that any ω-regular language[4] is measurable. On the other hand, not every measurable language is ω-regular: clearly any singleton $\{w\}$ generated by a word $w \in \Sigma^\omega$ is measurable, but the language $\{w\}$ may not be ω-regular if w is not a periodic word. It is also easy to construct an example of a non-measurable language: since Σ^ω is an uncountable Borel space, there is a Borel isomorphism $f : \Sigma^\omega \to [0,1]$, so for any non-Borel set $A \subseteq [0,1]$ the language $f^{-1}(A) \subseteq \Sigma^\omega$ is not measurable. We first show how we interpret languages as events in the canonical sample space H, and then introduce specific languages characterized by LTL formulae.

Consider a Borel measurable map $\mathsf{L} : X \to \Sigma$, henceforth called a *labelling* map. We call a triple $(\mathfrak{D}, \Sigma, \mathsf{L})$ a *labelled* `cdt`-MP (`lcdt`-MP for short): in a `lcdt`-MP each state $x \in X$ is assigned to a letter $\mathsf{L}(x) \in \Sigma$. As a result, to each path $h \in H$ there corresponds a unique *trace* word $w \in \Sigma^\omega$, also known as a *trace* of h, which is

[4] The definition of ω-regular languages is lengthy and is omitted from this paper for the sake of clarity in presentation. For a formal definition see e.g. [BK08, Section 4.3.1]

given by

$$\mathsf{L}_\omega(x_0, u_0, x_1, u_1, \dots) := (\mathsf{L}(x_0), \mathsf{L}(x_1), \dots). \tag{2.11}$$

We consider (2.11) as the definition of the *trace* map $\mathsf{L}_\omega : H \to \Sigma^\omega$.

Proposition 1. *The map L_ω is Borel measurable.*

Proof. Recall that $\mathscr{B}(\Sigma^\omega) = \sigma(\mathscr{C})$, where the class $\mathscr{C}$ of cylinder sets is given by

$$\mathscr{C} = \left\{ \prod_{k=0}^{n} C_k \times \prod_{k=n+1}^{\infty} \Sigma \,\middle|\, C_k \in \mathscr{B}(\Sigma), n \in \mathbb{N}_0 \right\}. \tag{2.12}$$

Next, for any cylinder set $C \in \mathscr{C}$, it holds that

$$\begin{aligned} \mathsf{L}_\omega^{-1}(C) &= \mathsf{L}_\omega^{-1}\left(\prod_{k=0}^{n} C_k \times \prod_{k=n+1}^{\infty} \Sigma \right) \\ &= \prod_{k=0}^{n} \left(\mathsf{L}^{-1}(C_k) \times U \right) \times \prod_{k=n+1}^{\infty} (X \times U) \in \mathscr{B}(H). \end{aligned}$$

From [Fol99, Proposition 2.1] it follows that L_ω is Borel measurable. □

It follows from Proposition 1 that given a `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$, for each measurable language $\phi \in \mathscr{B}(\Sigma^\omega)$ there corresponds a unique event $\mathsf{L}_\omega^{-1}(\phi) \in \mathscr{B}(H)$ that is the set of all paths of $\mathfrak{D}$ whose traces are elements of ϕ. In order to construct languages of interest in a natural way, we use LTL formulae. The grammar of LTL over the alphabet Σ is given by

$$\Phi \quad ::= \quad \sigma \in \Sigma \mid \neg\Phi \mid \Phi_1 \wedge \Phi_2 \mid \mathsf{X}\Phi \mid \Phi_1 \mathsf{U}^\infty \Phi_2. \tag{2.13}$$

The definition (2.13) shall be understood as follows: if Φ_1 and Φ_2 are LTL formulae, so are the expressions $\Phi_1 \wedge \Phi_2$, $\Phi_1 \mathsf{U}^\infty \Phi_2$, $\neg\Phi_1$ etc. Here $\wedge$ is the standard logical *conjunction* and $\neg$ is the logical *negation*, which allows us defining *disjunction* as $\Phi_1 \vee \Phi_2 := \neg(\neg\Phi_1 \wedge \neg\Phi_2)$. Furthermore, X and U^∞ are the *neXt* and *unbounded Until* temporal modalities whose meaning is clarified below.

The semantics of LTL formulae is defined using the notion of *accepted language*, that is $\mathfrak{L}(\Phi) \subseteq \Sigma^\omega$ is the collection of all infinite words over Σ that are accepted by the formula Φ. Firstly, we define the shift (or tail) on infinite words $\theta : \Sigma^\omega \to \Sigma^\omega$ by

$$\theta(w_0, w_1, w_2, \dots) = (w_1, w_2, \dots).$$

The semantics of LTL formulae is defined recursively as:

$$\begin{aligned} w \in \mathfrak{L}(\sigma) &\iff w_0 = \sigma \\ w \in \mathfrak{L}(\neg\Phi) &\iff w \notin \mathfrak{L}(\Phi) \\ w \in \mathfrak{L}(\Phi_1 \wedge \Phi_2) &\iff w \in \mathfrak{L}(\Phi_1) \cap \mathfrak{L}(\Phi_2) \\ w \in \mathfrak{L}(\mathsf{X}\Phi) &\iff \theta(w) \in \mathfrak{L}(\Phi), \end{aligned}$$

and in addition the semantics of the U^∞ modality is as follows:

$$w \in \mathfrak{L}(\Phi_1 \mathsf{U}^\infty \Phi_2) \iff \begin{aligned} &\theta^i(w) \in \mathfrak{L}(\Phi_2) \text{ for some } i \in \mathbb{N}_0 \text{ and} \\ &\theta^j(w) \in \mathfrak{L}(\Phi_1) \text{ for all } 0 \le j < i. \end{aligned} \tag{2.14}$$

It is useful to consider formulae describing bounded time horizon properties. We first introduce powers of X inductively as $\mathsf{X}^0\Phi := \Phi$ and $\mathsf{X}^n\Phi := \mathsf{X}(\mathsf{X}^{n-1}\Phi)$ for $n \ge 1$. Using the latter notation, it is now possible for any $n \in \mathbb{N}_0$ to define the formula

$$\Phi_1 \mathsf{U}^n \Phi_2 := \bigvee_{i=0}^{n} \left(\bigwedge_{j=0}^{i-1} \mathsf{X}^j \Phi_1 \wedge \mathsf{X}^i \Phi_2 \right), \tag{2.15}$$

whose semantics is a finite-horizon equivalent of (2.14), that is

$$w \in \mathfrak{L}(\Phi_1 \mathsf{U}^n \Phi_2) \iff \begin{aligned} &\theta^i(w) \in \mathfrak{L}(\Phi_2) \text{ for some } 0 \le i \le n \text{ and} \\ &\theta^j(w) \in \mathfrak{L}(\Phi_1) \text{ for all } 0 \le j < i. \end{aligned}$$

Note that U^∞ could be also expressed using (2.15), but the countably infinite number of operations of conjunction needed are not explicitly allowed in the syntax of LTL. We further denote $\texttt{true} := \bigvee_{\sigma \in \Sigma} \sigma$, and introduce new temporal modalities: *eventually*, $\Diamond^n \Phi := \texttt{true}\mathsf{U}^n\Phi$, and *always*, $\Box^n\Phi := \neg\Diamond^n\neg\Phi$, for all $n \in \bar{\mathbb{N}}_0$. We further simplify the notation as $\mathsf{U} := \mathsf{U}^\infty$, $\Diamond := \Diamond^\infty$ and $\Box := \Box^\infty$. Note that an accepted language of any LTL formula is ω-regular[5] [Wol81], and hence it is measurable, so that a formula in LTL is a valid way to describe events.

Let us provide some examples of how LTL formulae can be used to describe events of interest. We start with some basic formulae: let us consider a `cdt`-MP $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ and let $A, B \in \mathscr{B}(X)$ be two disjoint sets. We label them as A and B respectively, that is we introduce a labeling map $\mathsf{L} : X \to \Sigma$ where $\Sigma = \{A, B, \perp\}$ and

$$\mathsf{L}(x) = \begin{cases} A, & \text{if } x \in A, \\ B, & \text{if } x \in B, \\ \perp, & \text{otherwise.} \end{cases}$$

Then the event $\{\mathbf{x}_k \in A, k \ge 0\}$ can be expressed as $\Box A$, $\{\exists k \le n : \mathbf{x}_k \in B\}$ as $\Diamond^n B$, $\{\mathbf{x}_k \in B \text{ infinitely often }\}$ as $\Box\Diamond B$, $\{\exists k : \mathbf{x}_j \in A, j \ge k\}$ as $\Diamond\Box A$, and finally the event $\{\exists k \le n : \mathbf{x}_k \in B \text{ and } \mathbf{x}_j \in A, j < k\}$ can be expressed as $A\mathsf{U}^n B$.

As an additional example, recall the power network model from Section 2.2 and let S be the safe set, and G_1, G_2 be the preliminary target sets for each subnetwork, and G be the final target set. Define the alphabet $\Sigma = \{S, G_1, G_2, G, \perp\}$, where $\perp$ corresponds to the unsafe (failure) set, and let L be the obvious labeling map, e.g. $\mathsf{L}(x) = S$ if and only if $x \in S$. The first task of being within the safe energy levels can be characterized by the formula $\Box S$, whereas

$$S \wedge (S \,\mathsf{U}\, (G_1 \wedge (G_1 \,\mathsf{U}\, G))) \quad \vee \quad S \wedge (S \,\mathsf{U}\, (G_2 \wedge (G_2 \,\mathsf{U}\, G)))$$

[5] On the other hand, notice that there exist ω-regular languages that are not expressible by any LTL formula [Wol81].

is the desired formula for the second task. Indeed, in the first case only the word $SSSSS\ldots$ is accepted, which is produced exactly by those paths h that stay in S forever. Similarly, the second formula only accepts those words that eventually have the letter G following G_1 or G_2, and never contain letter $\bot$, so that the path representing energy levels never visits unsafe states and reaches the high energy level over the first subnetwork and then over the second, while still keeping the first level high, or vice-versa. Notice that in this second formula, which is to be interpreted within linear-time semantics, we are forcing trajectories to visit either of the $G_i, i = 1, 2$ before reaching G, whereas the similar requirement on S is set to prevent ever entering the unsafe set, but does not bound trajectories to first visit exclusively S: this will be clarified in the expression of the formula as an automaton, and in the choice of the sets (where $G_i \cap G = \emptyset$, and $(G_i \cup G) \subseteq S$) within the numerical case study.

For a given a `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$ we shall consider the expected utility criterion $\mathsf{M}^{\pi}\left(x; \mathsf{L}_{\omega}^{-1}(\mathfrak{L}(\Phi))\right)$ abbreviated by $\mathsf{M}^{\pi}(x; \Phi)$. The main problem can be now formulated as follows:

Problem 1. *Given a* `lcdt`*-MP* $(\mathfrak{D}, \Sigma, \mathsf{L})$*, an LTL formula* Φ *and a precision* $\varepsilon > 0$ *characterize* $\mathsf{M}^{*}(x; \Phi)$ *and compute its value with a given precision level* ε*.*

Note that if one is able to solve Problem 1, then one can also compute $\mathsf{M}_{*}(x; \Phi)$ for any LTL formula Φ thanks to the duality $\mathsf{M}_{*}(x; \Phi) = 1 - \mathsf{M}^{*}(x; \neg\Phi)$ – and likewise for ε-level computations.

2.4. Automata specifications

Above we have formulated the main problem we are focusing on in this paper, which requires computing extremal probabilities of events expressed as LTL formulae over infinite paths of a `cdt`-MP. Although LTL provides a succinct way to express events, for algorithmic purposes an equivalent *automata*-based perspective turns out to be more effective. Automata are transition systems with inputs over a finite alphabet and simple acceptance conditions [BK08, Chapter 4]. An input word is accepted by the automaton if a corresponding *run* of the automaton satisfies the acceptance condition. Before we introduce these concepts formally, let us mention that we follow the literature and only consider deterministic automata (those for which the current input and state uniquely determine the next state – cf. next definition), as they can be easily composed with `lcdt`-MP models.

Definition 4 (Transition system)**.** *Given an alphabet* Σ*, a deterministic* transition system *over* Σ *is a tuple* $\mathscr{T} = (Q, q^s, \Sigma, \mathtt{t})$ *where* Q *is a finite set,* $q^s \in Q$*, and* $\mathtt{t} : Q \times \Sigma \to Q$ *is some map. In our work all the transition systems are assumed to be deterministic*[6]*.*

[6] A non-deterministic transition system is one where $\mathtt{t} : Q \times \Sigma \to 2^Q$, that is given a current state of the system $q \in Q$ and an input letter $\sigma \in \Sigma$, the successor state $q' \in \mathtt{t}(q, \sigma)$ is not uniquely defined. As such, non-determinism here can be understood as set-valued dynamics, rather than as stochastic dynamics.

Given a transition system $\mathscr{T} = (Q, q^s, \Sigma, \mathsf{t})$ we say that Q is its state space, q^s is the initial condition, Σ is the input alphabet and t is the transition map. Any word $w \in \Sigma^\omega$ induces a run $r \in Q^\omega$ of $\mathscr{T}$ which is defined as follows: $r_0 = q^s$ and $r_{k+1} = \mathsf{t}(r_k, w_k)$ for any $k \in \mathbb{N}_0$. We can then introduce a map $\mathsf{t}_\omega : \Sigma^\omega \to Q^\omega$ that assigns to any input word the corresponding run. An ω-automaton is defined as follows:

Definition 5 (Automaton)**.** *A deterministic ω-automaton is a pair $\mathfrak{A} = (\mathscr{T}, \mathsf{Acc})$ consisting of a transition system $\mathscr{T} = (Q, q^s, \Sigma, \mathsf{t})$ together with an acceptance condition $\mathsf{Acc} \in \mathscr{B}(Q^\omega)$. We only consider automata on infinite words, so from now on we omit ω in "ω-automaton".*

An acceptance condition of an automaton indicates which runs are accepted by the automaton ($r \in \mathsf{Acc}$) and which are not ($r \notin \mathsf{Acc}$). Similarly, we can say that a word is accepted by a deterministic automaton if the corresponding run is accepted. The literature has considered several versions of acceptance conditions for ω-automata. In the context of this work the following three are the most important:

(DRA) a *deterministic Rabin automaton* is a tuple $\mathfrak{A} = (Q, q^s, \Sigma, \mathsf{t}, (F'_i, F''_i)_{i \in I})$ where $(Q, q^s, \Sigma, \mathsf{t})$ is a transition system, I is some finite index set, and $F'_i, F''_i \subseteq Q$ for any $i \in I$. A DRA accepts a run $r \in Q^\omega$ if there exists $i \in I$ such that r visits F'_i an infinite number of times and F''_i a finite number of times.

(DBA) a *deterministic Büchi automaton* is a special case of a DRA with I being a singleton and $F'' = \emptyset$, that is: a DBA is a tuple $\mathfrak{A} = (Q, q^s, \Sigma, \mathsf{t}, F)$ where $(Q, q^s, \Sigma, \mathsf{t})$ is a transition system and $F \subseteq Q$ is a set of final states. A DBA accepts a run $r \in Q^\omega$ if r visits F an infinite number of times.

(DFA) a *deterministic finite automaton* is a special case of a DBA[7] with all final states having self-loops ($\mathsf{t}(q, \sigma) = q$ for any $q \in F$, $\sigma \in \Sigma$), that is: a DFA is a tuple $\mathfrak{A} = (Q, q^s, \Sigma, \mathsf{t}, F)$ where $(Q, q^s, \Sigma, \mathsf{t})$ is a transition system and $F \subseteq Q$ is a set of final states. A DFA accepts a run $r \in Q^\omega$ if it visits F at least once[8].

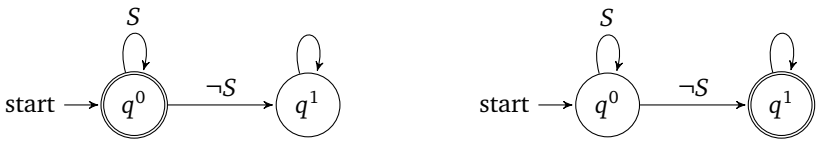
For an automaton $\mathfrak{A} = (\mathscr{T}, \mathsf{Acc})$ we define the accepted language of $\mathfrak{A}$ as the set of all infinite words that are accepted by $\mathfrak{A}$; we further denote this language by $\mathfrak{L}(\mathfrak{A})$, that is $\mathfrak{L}(\mathfrak{A}) := \mathsf{t}_\omega^{-1}(\mathsf{Acc})$. Similarly to Proposition 1, we can show that $\mathsf{t}_\omega \in \mathscr{B}(\Sigma^\omega)/\mathscr{B}(Q^\omega)$, so that $\mathfrak{L}(\mathfrak{A})$ is a measurable language as $\mathsf{Acc} \in \mathscr{B}(Q^\omega)$. Thus, for any `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$ with $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ the utility function $\mathsf{M}^\pi(\alpha; \mathsf{L}_\omega^{-1}(\mathfrak{L}(\mathfrak{A})))$ is well-defined. We further simplify the notation and write $\mathsf{M}^\pi(\alpha; \mathfrak{A})$.

[7] While it is canonical to introduce a DFA on finite words [BK08, Definition 4.9], we introduce it here on infinite words for the sake of consistency: in that way we do not have to consider both spaces of finite (Σ^*) and infinite (Σ^ω) words over the alphabet Σ, and can just focus on the latter. It shall be clear that our definition is also consistent with the canonical one in [BK08, Definition 4.9]: an infinite word $w \in \Sigma^\omega$ is accepted by a DFA if and only if there exists a finite prefix $w' \in \Sigma^*$ that is accepted by a DFA.

[8] An important version of the DFA has an n-horizon acceptance condition [TMKA13, Section 2.4], which requires the run to visit F in at most n steps. This is useful when one needs to express formulae in bounded LTL (BLTL) – a fragment of LTL (for details see Section 2.5).

Accepted languages of DRA are exactly ω-regular languages [BK08, Theorem 10.55], so in particular for any LTL formula Φ there exists a DRA $\mathfrak{A}^\Phi$ such that $\mathfrak{L}(\Phi) = \mathfrak{L}(\mathfrak{A}^\Phi)$. Furthermore, DBA (DFA) are strictly less expressive than DRA (DBA) – for details see [BK08, Chapter 4]. As further argued in the next section, rather than focusing on the most expressive DRA we consider all three kinds of automata. We will show that for any automaton $\mathfrak{A}$ the optimal utility $\mathsf{M}^*(x;\mathfrak{A})$ can be computed via a new optimal function $\hat{\mathsf{M}}^*((x,q^s);H \parallel \mathsf{Acc})$ over a newly defined `cdt`-MP $\hat{\mathfrak{D}}$, which is a composition of $\mathfrak{D}$ and $\mathfrak{A}$. Unfortunately, characterizing $\hat{\mathsf{M}}^*((x,q^s),H \parallel \mathsf{Acc})$ for DRA is rather difficult and we only provide partial results for the DBA case (Section 4), whereas the acceptance condition of the DFA allows for a much more complete characterization (Section 3).

Before we proceed, let us provide examples of automata for the tasks discussed over the power network model in Section 2.2. The DBA for the first task is given in Figure 2(a): here if we do not label the transition (as the loop at q^1) it means that the transition happens for any label. The final state is q^0 as indicated by a double circle. As we have mentioned above, the analysis of the DBA acceptance condition is more complicated than that of the DFA, hence even if the original LTL formula does not allow for the DFA expression, it is worth checking whether its negation does allow for one. For example, the DFA for the negation of the first task is given in Figure 2(b).



(a) DBA for the first task

(b) DFA for the negation of the first task

Figure 2: Automata representation of the first task of the case study in Section 2.2

The second task has a direct DFA expression, which is given in Figure 3. For an overview of available methods to construct an automaton given an LTL formula, see [VW94].

2.5. *Important fragments of LTL*

Although any LTL formula can be expressed as a DRA, such generality is not very useful in practice. Even when dealing with finite `cdt`-MP $\mathfrak{D}$, expressing a given formula as a DFA $\mathfrak{A} = (\mathscr{T}, D)$ (if possible) may reduce the complexity of the automaton, compared to some DRA expressions of the formula, as well as allows applying simpler solution methods, which altogether leads to a smaller state space of the composition $\mathfrak{D} \parallel \mathscr{T}$ and hence to a lower computational time. In the case when the `cdt`-MP $\mathfrak{D}$ is not finite, in addition the solution methods are much more involved and as Sections 3 and 4 will suggest, solution of a bounded-horizon reachability problem simpler than the one of an unbounded horizon reachability, which in turn is easier than the repeated reachability problem. As a result, e.g. although any LTL formula that encodes some bounded-horizon property can be expressed as

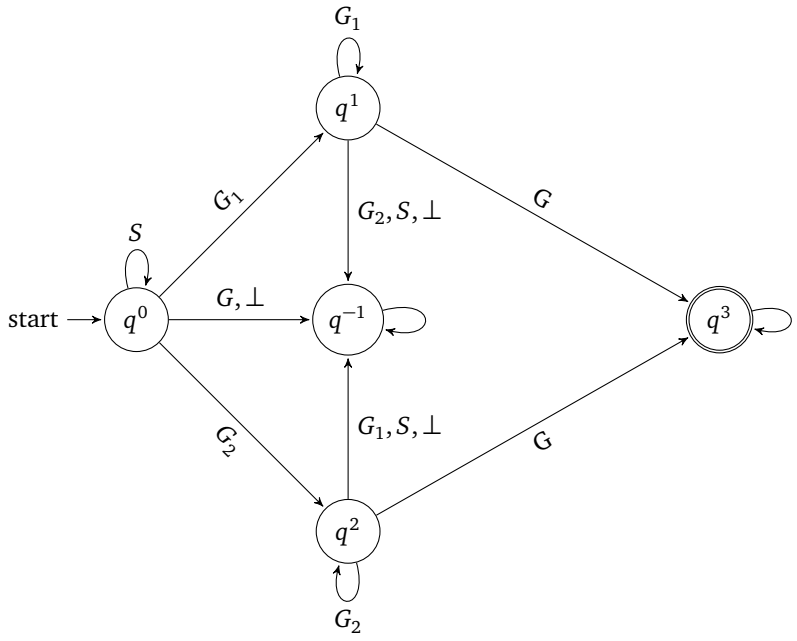


Figure 3: The DFA for the second task of the case study

a DRA, it is worth analyzing the formula to check whether it allows for an automaton expression with a simpler acceptance condition. In this section we describe how to perform such analysis, and what are the useful fragments of LTL that allow for an expression via an automaton that is simpler than a DRA.

The syntactically safe LTL (sLTL) [KV99] expresses safety languages. The language $\phi \subseteq \Sigma^\omega$ is called a safety property if and only if any word $w \notin \phi$ has a finite "bad" prefix:

$$w \notin \phi \quad \Longleftrightarrow \quad \exists n \in \mathbb{N}_0 : \mathsf{proj}^{-1}_{\Sigma^n}\left(\mathsf{proj}_{\Sigma^n}(w)\right) \cap \phi = \emptyset.$$

The syntactically co-safe LTL (scLTL) [KV99] expresses co-safety languages, where a co-safety language ϕ is the one for which any word $w \in \phi$ has a good prefix, that is

$$w \in \phi \quad \Longleftrightarrow \quad \exists n \in \mathbb{N}_0 : \mathsf{proj}^{-1}_{\Sigma^n}\left(\mathsf{proj}_{\Sigma^n}(w)\right) \subseteq \phi.$$

Clearly ϕ is a safety language if and only if $\Sigma^\omega \setminus \phi$ is a co-safety one. This comes as no surprise as safety languages are exactly closed subsets of Σ^ω in the product topology, whereas co-safety languages are open [AS85]. It follows that any co-safety language can be expressed as a DFA, and hence DFA can be used for negations of safety languages. Here we only provide a grammar for sLTL[9]. For this purpose, in

[9] The grammar of scLTL can be easily deduced from the one of sLTL; see also [AGLB12, Definition 2.1].

the LTL setting let us define a temporal modality *Weak until* W^∞ by

$$\Phi_1 \mathsf{W}^\infty \Phi_2 := \Phi_1 \mathsf{U} \Phi_2 \vee \Box \Phi_1.$$

The grammar of sLTL is given as follows:

$$\Phi \quad ::= \quad \sigma \in \Sigma \mid \neg\sigma \mid \Phi_1 \wedge \Phi_2 \mid \Phi_1 \vee \Phi_2 \mid \mathsf{X}\Phi \mid \Phi_1 \mathsf{W}^\infty \Phi_2.$$

Note that in sLTL the negation can be only applied on the level of letters, so that $\vee$ could not be expressed through $\wedge$ in general in sLTL in contrast to the LTL setting. Moreover, in general it is not possible to express $\Phi_1 \mathsf{U} \Phi_2$ using the sLTL grammar. An example of an sLTL formula is $\Box^n \sigma$, and that of an scLTL formula are $\Diamond^n \sigma$ and $\sigma_1 \mathsf{U}^n \sigma_2$where $n \in \mathbb{N}_0^\infty$ in all three cases. One immediate way to see whether a given LTL formula belongs to sLTL is to write it in a negation normal form (NNF), where the negation is presented on the level of atomic propositions by means of the following identities: $\neg \mathsf{X}\Phi = \mathsf{X}(\neg\Phi)$, $\neg(\Phi_1 \mathsf{U} \Phi_2) = \neg\Phi_1 \mathsf{W}^\infty \neg\Phi_2$ etc. However, even a LTL formula corresponding to a safety language may lead to a NNF which does not belong to sLTL, so for more elaborate methods see [KV99]. Recent examples of applications of sLTL and of scLTL can be found in [RMT13] and [AGLB12] respectively.

Although sLTL and scLTL are related to the expression of formulae via DFA rather than DRA, they still lead to the unbounded-horizon reachability problem over $\mathfrak{D} \parallel \mathscr{T}$, even in case when the original formula encodes a bounded-horizon specification. A useful framework to deal with the latter is given by the bounded LTL (BLTL) [TA13] which expresses bounded languages: a language $\phi \subseteq \Sigma^\omega$ is called bounded if there exists $n \in \mathbb{N}_0$ such that

$$w \in \phi \quad \Longleftrightarrow \quad \mathsf{proj}_{\Sigma^n}^{-1}\left(\mathsf{proj}_{\Sigma^n}(w)\right) \subseteq \phi.$$

In particular, it appears that bounded languages are exactly those that are both safety and co-safety languages [PP04, Proposition 3.10, Chapter III], that is they are clopen subsets of Σ^ω. The grammar of BLTL is given as follows:

$$\Phi \quad ::= \quad \sigma \in \Sigma \mid \neg\Phi \mid \Phi_1 \wedge \Phi_2 \mid \mathsf{X}\Phi, \tag{2.16}$$

so that it still allows for negations to be applied on all the levels, but U is absent from the syntax. On the other hand, (2.15) implies that $\Phi_1 \mathsf{U}^n \Phi_2$ belongs to BLTL for finite $n \in \mathbb{N}_0$. Any BLTL formula allows to be expressed as a bounded-horizon version of the DFA [TA13, Section 3.4] that accepts only those runs that visit the set of final states in at most n steps, where n is specified a priori, in the definition of the automaton. For applications of BLTL see e.g. [JCL+09].

2.6. Automata model checking

To state the main result of this section, we need to introduce the composition between an `lcdt`-MP and a transition system defined over the same alphabet, as formalised next.

Definition 6 (Composition of model and specification)**.** *Given an* `lcdt`*-MP* $(\mathfrak{D}, \Sigma, \mathsf{L})$ *with* $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ *and a transition system* $\mathscr{T} = (Q, q^s, \Sigma, \mathsf{t})$*, their* composition *is a* `cdt`*-MP* $\hat{\mathfrak{D}} = \mathfrak{D} \parallel \mathscr{T} = (\hat{X}, U, \hat{\mathbb{K}}, \hat{\mathsf{T}})$*, where* $\hat{X} := X \times Q$*,* $\hat{\mathbb{K}}_{(x,q)} := \mathbb{K}_x$ *for any* $x \in X$ *and* $q \in Q$ *and*

$$\hat{\mathsf{T}}(A \times B | x, q, u) := 1_B(\mathsf{t}(q, \mathsf{L}(x))) \cdot \mathsf{T}(A | x, u).$$

Let us further discuss this notion of composition. Consider an `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$ with $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ and a transition system $\mathscr{T} = (Q, q^s, \Sigma, \mathsf{t})$, and let $\hat{\mathfrak{D}} := \mathfrak{D} \parallel \mathscr{T}$. A more intuitive expression for the kernel $\hat{\mathsf{T}}$ can be given in the following way: given the current joint state (x_k, q_k) and control action u_k, the new state is

$$\begin{cases} x_{k+1} & \sim \mathsf{T}(x_k, u_k), \\ q_{k+1} & = \mathsf{t}(q_k, \mathsf{L}(x_k)). \end{cases}$$

The dynamics of the composed model should be understood as follows: the x-coordinate of the new state evolves according to the law T of the original `cdt`-MP $\mathfrak{D}$, and its label $\mathsf{L}(x)$ is used as an input to the transition system, which produces the q-coordinate. Let $\hat{H} := (\hat{X} \times U)^\omega$ denote the history space of $\hat{\mathfrak{D}}$, and further let $\hat{\Pi}$ be the class of all policies for $\hat{\mathfrak{D}}$ that give rise to strategic measures $\hat{\mathsf{P}}^{\hat{\pi}}_{\hat{\alpha}}$ for any $\hat{\pi} \in \hat{\Pi}$ and $\hat{\alpha} \in \mathscr{P}(\hat{X})$. We further let $(\hat{\mathbf{x}}_n)_{n \in \mathbb{N}_0} = (\mathbf{x}_n, \mathbf{q}_n)_{n \in \mathbb{N}_0}$, $(\mathbf{u}_n)_{n \in \mathbb{N}_0}$ and $(\hat{\mathbf{h}}_n)_{n \in \mathbb{N}_0}$ denote the state, action and information processes on the sample space $\hat{H}$, respectively.

As anticipated above, the main result of this section is as follows. For any `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$ and for any automaton $\mathfrak{A} = (\mathscr{T}, \mathsf{Acc})$, which may for example express an LTL formula Φ, it holds that $\mathsf{M}^*(x; \mathfrak{A}) = \hat{\mathsf{M}}^*((x, q^s); H \parallel \mathsf{Acc})$, where $\hat{\mathsf{M}}^*$ is the optimal utility functional over the composed `cdt`-MP $\hat{\mathfrak{D}} := \mathfrak{D} \parallel \mathscr{T}$. To obtain this result, we first need to establish a policy equivalence between optimal utilities over $\mathfrak{D}$ and $\hat{\mathfrak{D}}$. More precisely, we connect classes Π and $\hat{\Pi}$ as follows. The former class can be treated as a subclass of the latter, where policies do not depend on q-coordinates of $\hat{h}_n \in \hat{H}_n$, so we let $\mathfrak{i} : \Pi \to \hat{\Pi}$ denote the corresponding embedding map. Conversely, we introduce a projection map $\mathfrak{p} : \hat{\Pi} \to \Pi$ by the formula

$$(\mathfrak{p}\hat{\pi})_n(x_0, u_0, x_1, u_1, \dots, x_n) := \hat{\pi}_n(x_0, q_0, u_0, x_1, q_1, u_1, \dots, x_n, q_n),$$

where $q_0 = q^s$ and $q_{k+1} = \mathsf{t}(q_k, \mathsf{L}(x_k))$, for all $0 \leq k < n$.

Lemma 1. *For any* $\alpha \in \mathscr{P}(X)$*, and any policies* $\pi \in \Pi$ *and* $\hat{\pi} \in \hat{\Pi}$*, it holds that*

$$\mathsf{P}^\pi_\alpha(\mathsf{L}_\omega(h) \in \mathfrak{L}(\mathfrak{A})) = \hat{\mathsf{P}}^{\mathfrak{i}\pi}_{\alpha \otimes \delta_{q^s}}(H \parallel \mathsf{Acc}), \qquad \hat{\mathsf{P}}^{\hat{\pi}}_{\alpha \otimes \delta_{q^s}}(H \parallel \mathsf{Acc}) = \mathsf{P}^{\mathfrak{p}\hat{\pi}}_\alpha(\mathsf{L}_\omega(h) \in \mathfrak{L}(\mathfrak{A})).$$

Proof. Let us introduce a map $\beta : H \to \hat{H}$ as $\beta := \mathrm{id}_H \parallel (\mathsf{t}_\omega \circ \mathsf{L}_\omega)$, so that given a path $h \in H$ this map returns a path $\hat{h} = \beta(h) \in \hat{H}$ which has the same x- and u-coordinates, and the q-coordinates of which are obtained using the automaton transition map. As a result, for any $\alpha \in \mathscr{P}(X)$ and any $\pi \in \Pi$ it holds that

$$\mathsf{P}^\pi_\alpha(\mathsf{L}_\omega(h) \in \mathfrak{L}(\mathfrak{A})) = \mathsf{P}^\pi_\alpha((\mathsf{t}_\omega \circ \mathsf{L}_\omega)(h) \in \mathsf{Acc}) = (\beta_* \mathsf{P}^\pi_\alpha)(H \parallel \mathsf{Acc}).$$

Applying definitions of maps $\mathfrak{i}$ and $\mathfrak{p}$ immediately yields the desired result. □

Before we apply Lemma 1 to characterize the optimal utility of $\mathfrak{D}$ via that of $\hat{\mathfrak{D}}$, let us recall that DFA and DBA are not closed under negations[10], that is if we are able to express an LTL formula Φ as a DFA or DBA $\mathfrak{A}$, there may not exist such an expression for $\neg\Phi$. Due to this reason, in the next theorem we explicitly formulate both the maximization and the minimization problems, which allow us to apply the results both in cases of events expressed as DFA and DBA, and in cases when the complement of the event can be expressed in these automata classes. As above, let $\mathfrak{A} = (\mathscr{T}, \mathsf{Acc})$ be some automaton over an alphabet Σ, $(\mathfrak{D}, \Sigma, \mathsf{L})$ be any `lcdt`-MP and let $\hat{\mathfrak{D}} = \mathfrak{D} \parallel \mathscr{T}$.

Theorem 1. *The following equalities hold true:*

$$\mathsf{M}^*(\alpha; \mathfrak{A}) = \hat{\mathsf{M}}^*(\alpha \otimes \delta_{q^s}; H \parallel \mathsf{Acc}), \qquad \mathsf{M}_*(\alpha; \mathfrak{A}) = \hat{\mathsf{M}}_*(\alpha \otimes \delta_{q^s}; H \parallel \mathsf{Acc}).$$

Proof. The proof follows directly from Lemma 1 and Lemma 9 (cf. Section A.2). □

Let us discuss the importance of the result in Theorem 1. Suppose we are given an `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$ where Σ and L are used to distinguish the sets of interest, and a property expressed as a DFA or a DBA $\mathfrak{A}$ over the alphabet Σ. Such an expression may encode the LTL formula Φ for the desired property. Instead of having to compute the maximal probability $\mathsf{M}^*(\alpha; \mathfrak{A})$ directly, we can focus on an equivalent problem over $\hat{\mathfrak{D}} := \mathfrak{D} \parallel \mathscr{T}$, and focus on the property $\Diamond F$ in the case when $\mathfrak{A}$ is a DFA, or on the property $\square\Diamond F$ when $\mathfrak{A}$ is a DBA. We refer to the former property as *reachability* and to the latter as *repeated reachability*. The rest of the article is focused on the solution of both problems, so the coming results are applicable to classes of properties expressed as DFA and DBA, thanks to Theorem 1.

2.7. Further comments on models and problem formulation

The exposition of the model in this work is rather standard and is similar to that in [HLL96, Section 2.2]. However the presented model class is more general: for example we allow for a feasibility set $\mathbb{K}$ that is analytic, and for universally measurable policies. It can be shown that whenever the initial distribution $\alpha \in \mathscr{P}(X)$ is fixed, for a large class of performance criteria including all expected bounded utility cases it is sufficient to consider only analytically measurable deterministic policies depending exclusively on the history of state coordinates [Bla76]. Moreover, one can sufficiently deal with Borel measurable policies, provided they do exist. However, if one is interested in finding a policy that is optimal or ε-optimal for any initial distribution, it is more convenient to deal with the class of universally measurable policies: the latter is rich enough to assure the existence of policies for many interesting problems – see e.g. the discussion in [BS78, Section 1.2]. This class also possesses some nice closure properties in contrast to the class of analytically measurable policies: e.g. the composition of two universally measurable functions is

[10] Recall that here DFAs are interpreted over infinite words: when interpreted over finite words they are closed under negations.

again universally measurable, but the composition of analytically measurable functions may not be analytically measurable. Such closure properties are important to ensure the appropriate measurability of the performance criterion with respect to the initial state. More details on this topic can be found in [SB79].

It is worth mentioning that there is an alternative approach to sequential decision making in a stochastic environment, which is known as *gambling* [DS65]. While we cannot cover all the details on gambling models, we emphasise that this difference with the `cdt`-MP is mainly conceptual: if the current state is x, instead of first making a choice of a control action u and drawing a new state according to the distribution $\mathsf{T}(x,u)$, in gambling the agent is allowed to choose the distribution of the new state directly, from the set of available *gambles* Γ_x[11]. The set $\Gamma \subseteq X \times \mathscr{P}(X)$ is called the *gambling house*. On a methodological level, the difference between the `cdt`-MP and gambling is that the latter extensively uses stopping time-like methods to derive most of the results, whereas the former is more focused on techniques based on dynamic programming. Finally, differences between `cdt`-MP and gambling models can also be established on technical points, such as measurability properties of the strategies they deal with. First of all, initially the research on gambling theory has been done in the framework of finitely-additive probability measures [DS65]. Later, gambling models have also been considered in the σ-additive framework, which made it possible to compare them with `cdt`-MP: for example, [Bla76] showed the equivalence between some classes of `cdt`-MP and gambling models – this result also holds for the `cdt`-MP models that we consider in the present paper. Further gambling models have been used more recently, e.g. in [MPS91] and [MS96b].

Research on gambling has broadly looked into the optimization of probabilities of given events. For example, [MPS91] has obtained results for safety properties (that are clearly also applicable to reachability analysis), and [MPS91, MS96b] has characterized the repeated reachability property. Due to this reason, although we do not use the gambling framework explicitly, sometimes we recall the results obtained in this area. For example, using the `cdt`-MP framework for reachability properties seems more beneficial, however we mostly employ results of gambling for the repeated reachability property. Another important point is that [MS96a, Chapter 6] proposes an idea to optimize the probabilities of events, which is alternative to the one we convey in Section 2.3. More precisely, it is shown that in the case of a countable state space the functional M^* possesses some useful properties of the capacities [Del81]. In particular, [MS96a, Theorem (1.2), Chapter 6] claims that for any state $x \in X$ and any event $A \in \mathscr{B}(H)$ it holds that

$$\mathsf{M}^*(x;A) = \inf\{\mathsf{M}^*(x;B) : B \text{ is open and } B \supseteq A\}. \tag{2.17}$$

Furthermore, M^* for open events can be obtained by means of stopping times – see

[11] Note that in `cdt`-MP the choice of the distribution of the successor state is "labelled" by actions, whereas in gambling models such choice is unlabelled. One may think of this being similar to internal and external non-determinism in probabilistic automata [SL95], however there is no semantic difference between `cdt`-MP and gambling models, and in both cases non-determinism can be considered both as an internal one or as an external one.

[MS96a, Chapter 6] for more details. This result may be extensible to the more general case we deal with, where X is uncountable and one is interested only in events that can be described using some finite alphabet Σ. Unfortunately (2.17) does not provide a direct and explicit way to compute quantities of interest, or to derive optimal policies, so we do not pursue such direction here, preferring instead more explicit methods based on LTL formulae and automata theory.

The problem of optimizing the probability of a given event (or a property) is a problem that often appears in computer science, see e.g. a wide range of examples described in [BK08, Section 10.6]. Using LTL and automata theory for finite state-space `cdt`-MP has a long history, part of which can be consulted in [BK08, Section 10.8]. However, extensions to the general state-space case have only appeared recently: [AKM11] has provided an extension to the uncontrolled case (where trivially $U = \{u\}$ is a singleton), whereas [KSL13] and [TMKA13] worked out the controlled case[12]. In particular, the latter contribution is a basis for Section 2 and 3 of the current manuscript.

3. Reachability

3.1. Reachability problem: characterization

As Theorem 1 showed, optimizing probabilities over a `cdt`-MP for a large class of events of interest can be reduced to either a reachability problem, or to a repeated reachability one. This section is focused on the reachability problem. For this purpose it is more convenient to consider a slightly more general setup, called the *constrained reachability* problem [BK08, Section 10.1.1].[13] To satisfy the constrained reachability property, the path of a `cdt`-MP does not only have to reach a given goal set, but also to stay within some safe set before hitting the goal one. In terms of the LTL grammar, we are going to deal with the property $S\mathsf{U}^n G$ and a finite-time context, where S is a safe set and G is a goal set. The *(unconstrained)* reachability problem corresponds to the special case $\Diamond^n G = \mathtt{true}\mathsf{U}^n G$ (and likewise for the infinite-horizon case).

More precisely, consider a `cdt`-MP $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ and let $G \in \mathscr{B}(X)$ be the set of goal states, and $S \in \mathscr{B}(X)$ be the set of safe states. Define $D := S^c \setminus G$ to be the corresponding set of unsafe (or dangerous) states. For any initial distribution α, any policy $\pi \in \Pi$, and any time horizon $n \in \bar{\mathbb{N}}_0$ we are thus interested in the value of $\mathsf{M}^\pi(\alpha, S\mathsf{U}^n G)$. It is more convenient to focus on the initial distribution supported on single points and thus consider a function $\mathsf{M}^\pi(\cdot, S\mathsf{U}^n G) : X \to [0,1]$, extending the results to arbitrary initial distributions at a later stage. Clearly, $\mathsf{M}^\pi(\cdot; S\mathsf{U}^n G) \in$

[12] The difference between the approaches in these two works is that [KSL13] has allowed for Markov policies only, but clearly the policies over the composed system may depend on the state of the transition system: the map $\mathfrak{p}$ can map Markov policies to history-dependent ones. To cope with this issue, extended Markov policies have been proposed in [KSL13], namely policies that can depend also on an additional historical variable – the state of the transition system, which is a deterministic function of the `cdt`-MP state history.

[13] The constrained reachability problem is also known as the *reach-avoid* problem [SL10].

$\mathrm{b}\mathscr{U}(X)$ for any $\pi \in \Pi$ and $n \in \bar{\mathbb{N}}_0$. Moreover, the sequence $(\mathsf{M}^\pi(x; S\mathsf{U}^n G))_{n\in\mathbb{N}_0}$ is non-decreasing in n and furthermore for any fixed $x \in X$

$$\mathsf{M}^\pi(x; S\mathsf{U}G) = \lim_{n\to\infty} \mathsf{M}^\pi(x; S\mathsf{U}^n G). \tag{3.1}$$

Obviously, the unconstrained reachability defined in Section 2.6 is a special instance of constrained reachability in case the safe set is the whole state space, i.e. $S = X$.[14] Note that the solution of the problem is partially known:

$$\mathsf{M}^\pi(x; S\mathsf{U}^n G) = \begin{cases} 1, & \text{if } x \in G, \\ 0, & \text{if } x \in D \end{cases} \tag{3.2}$$

and, as a result, the constrained reachability problem needs to be solved only for states in $S \setminus G$. On the other hand, without loss of generality we can assume that sets S and G are disjoint: this follows from the fact that $\mathsf{M}^\pi(x; S\mathsf{U}^n G) = \mathsf{M}^\pi(x; (S \setminus G)\mathsf{U}^n G)$. Below this assumption is often made for the sake of notation; this also allows us to highlight that the dynamics of `cdt`-MP over the set S are of the highest importance for the solution of the problem, in contrast to the dynamics of the states in the set G. As we have mentioned above, we consider both the maximization and the minimization problems for constrained reachability, namely both $\mathsf{M}^*(x; S\mathsf{U}^n G)$ and $\mathsf{M}_*(x; S\mathsf{U}^n G)$.

It is known that the DP principles allow decomposing the general optimization problem into smaller and simpler subproblems [Bel57]. In the literature there have been several results developing DP characterizations of the constrained reachability problem. One of the main differences in these studies has been the choice of the structural representation of the value function $\mathsf{M}(\cdot; S\mathsf{U}^n G)$. For example, the work in [APLS08] has considered the max cost representation for unconstrained reachability, as

$$\mathsf{M}^\pi(x; X\mathsf{U}^n G) = \mathsf{P}_x^\pi \left[\max_{k\le n} 1_G(\mathbf{x}_k) \right], \tag{3.3}$$

and using the dual safety problem, an alternative multiplicative cost representation

$$\mathsf{M}^\pi(x; X\mathsf{U}^n G) = 1 - \mathsf{P}_x^\pi \left[\prod_{k=0}^{n} 1_{G^c}(\mathbf{x}_k) \right]. \tag{3.4}$$

These results have been extended in [SL10], which has dealt with the general constrained reachability problem in the form of a sum-multiplicative cost

$$\mathsf{M}^\pi(x; S\mathsf{U}^n G) = \mathsf{P}_x^\pi \left[\sum_{k=0}^{n} \left(\prod_{j=0}^{k-1} 1_{S\setminus G}(\mathbf{x}_j) \right) 1_G(\mathbf{x}_k) \right]. \tag{3.5}$$

[14] As a side note, constrained reachability can be also obtained from the unconstrained one by changing the dynamics of the `cdt`-MP on the set D [TMKA13, Section 3.1].

Later, [CCL11] suggested a cost formulation using the notion of a first hitting time as

$$\mathsf{M}^{\pi}(x; S\mathsf{U}^{n}G) = \mathsf{P}_{x}^{\pi}\left[\sum_{k=0}^{n\wedge\tau_{G}\wedge\tau_{D}} 1_{G}(\mathbf{x}_{k})\right], \tag{3.6}$$

where $\tau_A := \inf\{k \geq 0 : \mathbf{x}_k \in A\}$ is the first hitting time of the set $A \in \mathscr{B}(X)$, and where $\wedge$ denotes the min operator. As we have mentioned in Section 2.3, the TC performance criterion allows for a rich theory of DP in a general setting. The aforementioned studies in [APLS08], [SL10] and [CCL11] have recovered only a subset of these results for the reachability problem, sometimes requiring restrictive assumptions on the structure of the model. Here we show that the reachability problem has an equivalent TC formulation, which allows us proving general results for this performance criterion.

In general it may not be possible to characterize the constrained reachability problem as a TC criterion over the original `cdt`-MP $\mathfrak{D}$. The key idea is to consider an auxiliary `cdt`-MP $\hat{\mathfrak{D}}$, constructed from the original one by adding a new Boolean variable (taking value in the set $\{q^s, q^f\}$) that represents whether the path of $\mathfrak{D}$ has left the safe set S or not. To our knowledge, the first time such construction has been explicitly used in [TMKA13].[15] For the sake of consistency, here we introduce a new `cdt`-MP using the notion of the composition between the transition system and the original `cdt`-MP.

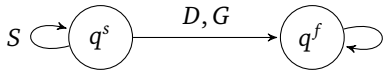


Figure 4: Transition system for the TC formulation of constrained reachability

Let us consider a transition system $\mathscr{T} = (Q, q^s, \Sigma, \mathsf{t})$ as in Figure 4 with a state space $Q = \{q^s, q^f\}$, an alphabet $\Sigma = (D, G, S)$, and a transition function given by

$$\mathsf{t}(q^s, S) = q^s, \quad \mathsf{t}(q^s, \{D, G\}) = q^f, \quad \mathsf{t}(q^f, \Sigma) = q^f.$$

We extend $\mathfrak{D}$ to a `lcdt`-MP $(\mathfrak{D}, \Sigma, \mathsf{L})$ trivially by letting $\mathsf{L}(x) = S$ for all $x \in S$ (the label of the states in the complement being irrelevant), and denote by $\hat{\mathfrak{D}} := (\mathfrak{D}, \Sigma, \mathsf{L}) \parallel \mathscr{T} = (\hat{X}, U, \hat{\mathbb{K}}, \hat{T})$ the composed `cdt`-MP. We also let the corresponding canonical probability space and related state, action and information processes to be defined as in Section 2.6. Let us explicitly write down the relation between operators $\hat{\mathsf{T}}$ and T, as they are further needed below:

$$\begin{aligned}\hat{\mathsf{T}}^{u}\hat{f}(x,q) &= 1\{q = q^f\}\mathsf{T}^{u}\hat{f}(x,q^f) \\ &\quad + 1\{q = q^s\}\left(1_S(x)\mathsf{T}^{u}\hat{f}(x,q^s) + 1_{S^c}(x)\mathsf{T}^{u}\hat{f}(x,q^f)\right),\end{aligned} \tag{3.7}$$

which holds for any $\hat{f} \in \mathrm{b}\mathscr{U}(\hat{X})$, $x \in X$, $u \in U$, and $q \in Q$. In particular, of special interest are functions $\hat{f} : \hat{X} \to \mathbb{R}$ that are zero off q^s, namely $\hat{f}(\cdot, q^f) \equiv 0$: they can

[15] In [DAT13] a similar construction was used to formulate the reachability problem as a final cost problem.

be always represented in the form

$$\hat{f}(x,q) = 1\{q = q^s\} \cdot f(x) \tag{3.8}$$

for some $f : X \to \mathbb{R}$. For functions as in (3.8), equation (3.7) simplifies as

$$\hat{\mathsf{T}}^u \hat{f}(x,q) = 1\{q = q^s\} \mathsf{T}^u \hat{f}(x,q^s) = 1\{q = q^s\} \mathsf{T}^u f(x) \tag{3.9}$$

so that $\hat{\mathsf{T}}^u$ preserves the property of being zero off q^s.

Let $c : \hat{X} \to \{0,1\}$ be a cost function $c(x,q) := 1\{q = q^s\} \cdot 1_G(x)$ that is zero off q^s, and define the corresponding TC utility for any $n \in \bar{\mathbb{N}}_0$ as follows:

$$\hat{J}_n := \sum_{k=0}^{n} c(\mathbf{x}_k, \mathbf{q}_k), \tag{3.10}$$

where $\mathbf{x}$ and $\mathbf{q}$ are the components of the state process as in Section 2.6. The corresponding maximization and minimization problems are given by

$$\hat{\mathsf{M}}^*(x,q;\hat{J}_n) = \sup_{\hat{\pi} \in \hat{\Pi}} \hat{\mathsf{P}}^{\pi}_{(x,q)}\left[\hat{J}_n\right], \qquad \hat{\mathsf{M}}_*(x,q;\hat{J}_n) = \inf_{\hat{\pi} \in \hat{\Pi}} \hat{\mathsf{P}}^{\pi}_{(x,q)}\left[\hat{J}_n\right]. \tag{3.11}$$

In order to show the equivalence between the optimal constrained reachability problem over the `cdt`-MP $\mathfrak{D}$ and the formulation in (3.11) over the `cdt`-MP $\hat{\mathfrak{D}}$, we apply the technique from Section 2.6. Let us denote by $\mathfrak{i} : \Pi \to \hat{\Pi}$ the obvious embedding map, and let the projection map $\hat{\mathfrak{p}} : \hat{\Pi} \to \Pi$ be given by

$$(\hat{\mathfrak{p}}\hat{\pi})_n(x_0, u_0, \ldots, x_{n-1}, u_{n-1}, x_n) := \hat{\pi}_n(x_0, q^s, u_0, \ldots, x_{n-1}, q^s, u_{n-1}, x_n, q^s). \tag{3.12}$$

Note that $\hat{\mathfrak{p}}$ is different from the projection map $\mathfrak{p}$ discussed in Section 2.6: in particular, later we use the fact that $\hat{\mathfrak{p}}(\hat{\Pi}_M) \subseteq \Pi_M$, whereas $\mathfrak{p}$ does not necessarily preserve the Markovian property of a policy. The following equivalence holds true:

Lemma 2. *For any $n \in \bar{\mathbb{N}}_0$, $\pi \in \Pi$ and $\hat{\pi} \in \hat{\Pi}$, it holds that*

$$\hat{\mathsf{M}}^{\hat{\pi}}(x,q^s;\hat{J}_n) = \mathsf{M}^{\hat{\mathfrak{p}}\hat{\pi}}(x; S\mathsf{U}^n G), \qquad \mathsf{M}^{\pi}(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}^{\mathfrak{i}\pi}(x,q^s;\hat{J}_n). \tag{3.13}$$

Proof. We prove this theorem by induction. First of all, both equalities in (3.13) clearly hold true for $n = 0$ as in this case all functions are simply $1_G(x)$. Furthermore, with focus on the first equality, we have that

$$\hat{\mathsf{M}}^{\hat{\pi}}(x,q^s;\hat{J}_{n+1}) - \hat{\mathsf{M}}^{\hat{\pi}}(x,q^s;\hat{J}_n) = \hat{\mathsf{P}}^{\hat{\pi}}_{(x,q^s)}\left[c\left(\mathbf{x}_{n+1}, \mathbf{q}_{n+1}\right)\right].$$

As $c(\mathbf{x}_{n+1}, \mathbf{q}_{n+1})$ is a Bernoulli random variable supported on $\{0,1\}$, we obtain that

$$\begin{aligned}\hat{\mathsf{P}}^{\hat{\pi}}_{(x,q^s)}\left[c\left(\mathbf{x}_{n+1}, \mathbf{q}_{n+1}\right)\right] &= \hat{\mathsf{P}}^{\hat{\pi}}_{(x,q^s)}\left(c\left(\mathbf{x}_{n+1}, \mathbf{q}_{n+1}\right) = 1\right) \\ &= \hat{\mathsf{P}}^{\hat{\pi}}_{(x,q^s)}\left(\{\mathbf{x}_k \in S, k \le n\}, \{\mathbf{x}_{n+1} \in G\}, \{\mathbf{q}_k = q^s, k \le n+1\}\right).\end{aligned}$$

On the other hand, the increment in n of the function $\mathsf{M}^{\hat{\mathfrak{p}}\hat{\pi}}$ is

$$\mathsf{M}^{\hat{\mathfrak{p}}\hat{\pi}}(x; S\mathsf{U}^{n+1} G) - \mathsf{M}^{\hat{\mathfrak{p}}\hat{\pi}}(x; S\mathsf{U}^n G) = \mathsf{P}^{\hat{\mathfrak{p}}\hat{\pi}}_x\left(\{\mathbf{x}_k \in S, k \le n\}, \{\mathbf{x}_{n+1} \in G\}\right).$$

The fact that these probabilities are equal follows immediately from their integral expressions in (2.6) and from the definition of the projection map $\hat{\mathfrak{p}}$. By induction we obtain the first part in (3.13) for $n < \infty$, and the case $n = \infty$ follows by taking the limit. Finally, the proof of the second part of (3.13) is obtained the same way, mutatis mutandis. □

Lemma 2 leads to several important results that allow us to develop a DP framework for constrained reachability. First of all, it clearly implies that both optimization problems are equivalent in the following sense:

Theorem 2. *For all $n \in \bar{\mathbb{N}}_0$ and $x \in X$ we have $\hat{\mathsf{M}}^*(x, q^f; \hat{J}_n) = \hat{\mathsf{M}}_*(x, q^f; \hat{J}_n) = 0$ and*

$$\mathsf{M}^*(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}^*(x, q^s; \hat{J}_n), \quad \mathsf{M}_*(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}_*(x, q^s; \hat{J}_n). \tag{3.14}$$

Proof. To prove the first part, one has to notice that if $\mathbf{q}_0 = q^f$, then $\mathbf{q}_n = q^f$ for all $n \in \mathbb{N}_0$, hence $\hat{\mathsf{M}}^{\hat{\pi}}(x, q^f; \hat{J}_n) = 0$ for all $n \in \mathbb{N}_0$, $x \in X$, and $\hat{\pi} \in \hat{\Pi}$. Furthermore, (3.14) is an immediate consequence of Lemma 2 and Lemma 9 in the Appendix. □

As we have mentioned above, Theorem 2 allows us to extrapolate the rich theory developed for the TC criterion to the case of the constrained reachability problem. However, most of the results for TC are developed for the minimization case [BS78, HLL96], considering either positive or negative costs c. As such, we can directly derive the results for the minimization problem since $\mathsf{M}_*(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}_*(x, q^s; \hat{J}_n)$, however for the maximal constrained reachability we shall interpret

$$\mathsf{M}^*(x; S\mathsf{U}^n G) = -\hat{\mathsf{M}}_*(x, q^s; -\hat{J}_n),$$

thus characterizing both optimization problems as a minimization of some TC. Note that for the minimization of constrained reachability we use a positive cost c, thus falling into the setting of the positive DP [Bla76] corresponding to [BS78, Assumption (P), Chapter 9]. On the other hand, for the maximization of constrained reachability a negative cost $-c$ is used, hence leading to the case of the negative DP [Str66] corresponding to [BS78, Assumption (N), Chapter 9]. This difference is not always important and only matters in the case $n = \infty$, however we show below that we need to distinguish the two cases to prove the convergence of bounded-horizon functions to the unbounded-horizon ones, as well as the existence of optimal policies.

Let us proceed with the application of Lemma 2 and Theorem 2 to the characterization of the optimal constrained reachability problems. The next result shows that it is sufficient to deal with Markov policies.

Proposition 2. *For any $n \in \bar{\mathbb{N}}_0$ and any policy $\pi \in \Pi$, there exists a Markov policy $\pi' \in \Pi_M$ such that $\mathsf{M}^{\pi}(\cdot; S\mathsf{U}^n G) = \mathsf{M}^{\pi'}(\cdot; S\mathsf{U}^n G)$, and as a consequence*

$$\mathsf{M}^*(x; S\mathsf{U}^n G) = \sup_{\pi \in \Pi_M} \mathsf{M}^{\pi}(x; S\mathsf{U}^n G), \quad \mathsf{M}_*(x; S\mathsf{U}^n G) = \inf_{\pi \in \Pi_M} \mathsf{M}^{\pi}(x; S\mathsf{U}^n G). \tag{3.15}$$

Proof. Fix any state $x \in X$ and any policy $\pi \in \Pi$. It follows from Lemma 2 that $\mathsf{M}^{\pi}(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}^{\mathsf{i}\pi}(x, q^s; \hat{J}_n)$. On the other hand, [BS78, Proposition 8.1] assures the existence of a Markov policy $\hat{\pi}' \in \hat{\Pi}_M$ satisfying $\hat{\mathsf{M}}^{\mathsf{i}\pi}(x, q^s; \hat{J}_n) = \hat{\mathsf{M}}^{\hat{\pi}'}(x, q^s; \hat{J}_n)$. From the definition of the projection map $\hat{\mathfrak{p}}$ it follows that $\pi' := \hat{\mathfrak{p}}\hat{\pi}' \in \Pi_M$ and as a result

$$\mathsf{M}^{\pi'}(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}^{\hat{\pi}'}(x, q^s; \hat{J}_n) = \hat{\mathsf{M}}^{\mathsf{i}\pi}(x, q^s; \hat{J}_n) = \mathsf{M}^{\pi}(x; S\mathsf{U}^n G),$$

as desired. In order to obtain (3.15) we only have to apply Lemma 9. □

The results above, obtained for deterministic initial conditions, can be extended to the case of general initial distributions: we show that a value function over an initial distribution $\alpha \in \mathscr{P}(X)$ can be obtained by integrating value functions over deterministic initial conditions. Although this result is obvious whenever the policy is fixed, it is not trivial to be shown for optimal value functions. We show a proof for the case of the minimization problem on the unbounded time horizon, however similar results can be obtained for the unbounded-time maximization case, as well as for both bounded-horizon problems.

Proposition 3. *For any distribution $\alpha \in \mathscr{P}(X)$ it holds that*

$$\mathsf{M}_*(\alpha; S\mathsf{U}G) = \int_X \mathsf{M}_*(x; S\mathsf{U}G)\, \alpha(\mathrm{d}x). \tag{3.16}$$

Proof. From [BS78, Propositions 9.2, 9.3, 9.5] it follows that

$$\hat{\mathsf{M}}_*(\hat{\alpha}; \hat{J}_\infty) = \int_{\hat{X}} \hat{\mathsf{M}}_*(x, q; \hat{J}_\infty) \hat{\alpha}(\mathrm{d}x \times \mathrm{d}q)$$

for any distribution $\hat{\alpha} \in \mathscr{P}(\hat{X})$. As a result, for any $\alpha \in \mathscr{P}(X)$ it holds that

$$\begin{aligned}
\mathsf{M}_*(\alpha; S\mathsf{U}G) &= \inf_{\pi \in \Pi} \int_{\hat{X}} \hat{\mathsf{M}}^{\mathsf{i}\pi}(x, q; \hat{J}_\infty)(\alpha \otimes \delta_{q^s})(\mathrm{d}x \times \mathrm{d}q) \\
&\geq \int_{\hat{X}} \hat{\mathsf{M}}_*(x, q; \hat{J}_\infty)(\alpha \otimes \delta_{q^s})(\mathrm{d}x \times \mathrm{d}q) \\
&= \int_X \hat{\mathsf{M}}_*(x, q^s; \hat{J}_\infty)\alpha(\mathrm{d}x) = \int_X \mathsf{M}_*(x; S\mathsf{U}G)\, \alpha(\mathrm{d}x).
\end{aligned}$$

The converse inequality we get as follows:

$$\begin{aligned}
\int_X \mathsf{M}_*(x; S\mathsf{U}G)\, \alpha(\mathrm{d}x) &= \int_X \hat{\mathsf{M}}_*(x, q^s; \hat{J}_\infty)\alpha(\mathrm{d}x) \\
&= \int_{\hat{X}} \hat{\mathsf{M}}_*(x, q; \hat{J}_\infty)(\alpha \otimes \delta_{q^s})(\mathrm{d}x \times \mathrm{d}q) \\
&= \inf_{\hat{\pi} \in \hat{\Pi}} \int_{\hat{X}} \hat{\mathsf{M}}^{\hat{\pi}}(x, q; \hat{J}_\infty)(\alpha \otimes \delta_{q^s})(\mathrm{d}x \times \mathrm{d}q) \\
&= \inf_{\hat{\pi} \in \hat{\Pi}} \int_X \mathsf{M}^{\hat{\mathfrak{p}}\hat{\pi}}(x; S\mathsf{U}G)\alpha(\mathrm{d}x) \geq \mathsf{M}_*(\alpha; S\mathsf{U}G).
\end{aligned}$$

Since both inequalities hold true, we obtain the desired result. □

Although in general one cannot switch the order of a minimization and an integral, Proposition 3 shows this can be done in the case of (3.16). As a consequence, for the sake of simplicity we can deal directly with deterministic initial distributions: the value function for general ones can be obtained by integrating with respect to the initial distribution of interest.

We are ready to formulate one of the most relevant outcomes of Theorem 2: a DP procedure for the constrained reachability problem over a general class of policies. For this purpose we introduce the following DP operators:

$$\begin{aligned} \mathsf{R}^* f(x) &= 1_G(x) + 1_S(x) \cdot \mathsf{T}^* f(x), \qquad f \in \mathrm{b}\mathscr{A}^*(X), \\ \mathsf{R}_* f(x) &= 1_G(x) + 1_S(x) \cdot \mathsf{T}_* f(x), \qquad f \in \mathrm{b}\mathscr{A}_*(X). \end{aligned}$$

From the properties of operators T^* and T_*, it follows that R^* maps $\mathrm{b}\mathscr{A}^*(X)$ into itself and R_* maps $\mathrm{b}\mathscr{A}_*(X)$ into itself.

Theorem 3. *It holds that* $\mathsf{M}^*(\cdot; S\mathsf{U}^0 G) = \mathsf{M}_*(\cdot; S\mathsf{U}^0 G) = 1_G(\cdot)$*, and for any* $n \in \bar{\mathbb{N}}_0$

$$\mathsf{M}^*(\cdot; S\mathsf{U}^{n+1} G) = \mathsf{R}^*\left[\mathsf{M}^*(\cdot; S\mathsf{U}^n G)\right], \qquad \mathsf{M}_*(\cdot; S\mathsf{U}^{n+1} G) = \mathsf{R}_*\left[\mathsf{M}_*(\cdot; S\mathsf{U}^n G)\right].$$

Moreover, $\mathsf{M}^*(\cdot; S\mathsf{U}G)$ *and* $\mathsf{M}_*(\cdot; S\mathsf{U}G)$ *are the least non-negative fixpoints of the corresponding operators, that is if there exists a non-negative function* $f \in \mathrm{b}\mathscr{A}^*(X)$ *(or* $f \in \mathrm{b}\mathscr{A}_*(X)$*) that satisfies the inequality* $f \geq \mathsf{R}^*[f]$ *(or* $f \geq \mathsf{R}_*[f]$*), then it holds that* $f(\cdot) \geq \mathsf{M}^*(\cdot; S\mathsf{U}G)$ *(or* $f(\cdot) \geq \mathsf{M}_*(\cdot; S\mathsf{U}G)$*).*

Proof. We provide an explicit proof for the minimization problem, and appeal to duality for the maximization case.

First of all, the fact that $\mathsf{M}_*(\cdot; S\mathsf{U}^0 G) = 1_G(\cdot)$ follows immediately from the definition of constrained reachability. Furthermore, for any $n \in \bar{\mathbb{N}}_0$ by Theorem 2 we have that $\mathsf{M}_*(x; S\mathsf{U}^n G) = \hat{\mathsf{M}}_*(x, q^s; \hat{J}_n)$. The DP recursion for the TC is given in [BS78, Proposition 8.2, Proposition 9.8], and applied here yields the following:

$$\begin{aligned} \mathsf{M}_*(x; S\mathsf{U}^{n+1} G) = \hat{\mathsf{M}}_*(x, q^s; \hat{J}_{n+1}) &= \inf_{u \in \mathbb{K}_x} \left(c(x, q^s) + \hat{\mathsf{T}}^u \left[\hat{\mathsf{M}}_*(x, q^s; \hat{J}_n) \right] \right) \\ &= \inf_{u \in \mathbb{K}_x} \left(1_G(x) + 1_S(x) \mathsf{T}^u \left[\hat{\mathsf{M}}_*(x, q^s; \hat{J}_n) \right] \right) \\ &= 1_G(x) + 1_S(x) \mathsf{T}_* \left[\mathsf{M}_*(x; S\mathsf{U}^n G) \right] = \mathsf{R}_* \left[\mathsf{M}_*(x; S\mathsf{U}^n G) \right]. \end{aligned}$$

This results in both the DP recursion ($n < \infty$) and in the fixpoint equation (for $n = \infty$).

Consider now a non-negative function $f \in \mathrm{b}\mathscr{A}_*(X)$ satisfying $f \geq \mathsf{R}_*[f]$, and define a new function $\hat{f} : \hat{X} \to [0, \infty)$ by the formula $\hat{f}(x, q) := 1\{q = q^s\} \cdot f(x)$. Clearly, the function $\hat{f}$ is zero off q^s, so that we obtain:

$$\inf_{u \in \mathbb{K}_x} \left(c(x, q) + \hat{\mathsf{T}}^u \hat{f}(x, q) \right) = 1\{q = q^s\} \cdot \mathsf{R}_* f(x) \leq 1\{q = q^s\} \cdot f(x) = \hat{f}(x, q).$$

As a result, [BS78, Proposition 9.10 (P)] implies that $\hat{\mathsf{M}}_*(\cdot; J_\infty) \leq \hat{f}(\cdot)$ and thus

$$\mathsf{M}_*(x; S\mathsf{U}G) = \hat{\mathsf{M}}_*(x, q^s; \hat{J}_\infty) \leq \hat{f}(x, q^s) = f(x),$$

so $\mathsf{M}_*(\cdot; S\mathsf{U}G)$ is the least fixpoint in the class of non-negative $\mathrm{b}\mathscr{A}_*$ functions. □

In view of Theorem 3 we can compute the value of the bounded horizon optimal constrained reachability problems backward-recursively, starting from the indicator function 1_G. The computation of the fixpoint problem is more intricate and is addressed below in Section 3.2. Due to this reason, it is worth discussing the relation between the solution of the constrained reachability problem on the bounded time horizon, and that on the unbounded time horizon. In particular, an interesting question is whether the latter can be in general obtained as the limit of the former, as the time index n goes to infinity. This is one of the anticipated cases where the difference between the maximization and minimization problems becomes important: the answer is positive in the first case and is negative in the second.

Proposition 4. *For every state $x \in X$ it holds that*

$$\mathsf{M}^*(x; S\mathsf{U}G) = \lim_{n\to\infty} \mathsf{M}^*(x; S\mathsf{U}^n G). \tag{3.17}$$

Furthermore, for any $x \in X$ there exists a limit

$$f_*(x) := \lim_{n\to\infty} \mathsf{M}_*(x; S\mathsf{U}^n G) \le \mathsf{M}_*(x; S\mathsf{U}G). \tag{3.18}$$

Moreover, $\mathsf{M}_*(\cdot; S\mathsf{U}G) = f_*(\cdot)$ if and only if f_* is a fixpoint of the DP operator R_*.

Proof. We start with the maximization case: recall that it corresponds to Assumption (N) of [BS78, Chapter 9] since $\mathsf{M}^*(x; S\mathsf{U}^n G) = -\hat{\mathsf{M}}_*(x, q^s; -\hat{J}_n)$ for any $x \in X$. It follows from [BS78, Section 9.5] that the sequence $(\hat{\mathsf{M}}_*(x, q; -\hat{J}_n))_{n\in\mathbb{N}}$ has a limit for any $x \in X$ and $q \in Q$. Furthermore, [BS78, Proposition 9.14] implies that this limit is $\hat{\mathsf{M}}_*(x, q; -\hat{J}_\infty)$, which leads to (3.17).

For the minimization case we satisfy Assumption (P) of [BS78, Chapter 9]. The discussion in [BS78, Section 9.5] implies the existence of the point-wise limit for the sequence $(\hat{\mathsf{M}}_*(x, q; \hat{J}_n))_{n\in\mathbb{N}}$: we denote this limit by $\hat{f}_*$. Furthermore, it follows from [BS78, Proposition 9.16] that $\hat{f}_*(\cdot) \le \hat{\mathsf{M}}_*(\cdot; \hat{J}_\infty)$, and that the equality holds if and only if $\hat{f}_*$ is a fixpoint of the corresponding DP operator, i.e.

$$\hat{f}_*(x, q) = c(x, q) + \hat{\mathsf{T}}_* \hat{f}_*(x, q). \tag{3.19}$$

For the constrained reachability case, we now obviously have the existence of the limit

$$f_*(x) := \lim_{n\to\infty} \mathsf{M}_*(x; S\mathsf{U}^n G) = 1\{q = q^s\} \hat{f}_*(x, q).$$

Clearly, $f_*(\cdot) \ge \mathsf{M}_*(\cdot; S\mathsf{U}G)$. Furthermore, if f_* is a fixpoint of R_*, then $\hat{f}_*$ satisfies (3.19), thus $\hat{f}_*(\cdot) = \hat{\mathsf{M}}(\cdot; \hat{J}_\infty)$ and $f_*(\cdot) = \mathsf{M}_*(\cdot; S\mathsf{U}G)$. Conversely, if $f_*(\cdot) = \mathsf{M}_*(\cdot; S\mathsf{U}G)$ then by Theorem 3 it has to be a fixpoint of the DP operator R_*. □

The following example shows that the inequality in (3.18) can be strict.[16]

[16] The example is obtained by modifying [BS78, Example 1].

Example 1. *Let $X = \mathbb{N}_0$ and let $U = \{1/n\}_{n\in\mathbb{N}} \cup \{-1\}$. Define admissible controls as follows: $\mathbb{K}_0 = \{1/n\}_{n\in\mathbb{N}}$ and $\mathbb{K}_x = -1$ for $x \neq 0$. The dynamics is deterministic and is given by the following update law:*

$$\mathbf{x}_{n+1} = \mathbf{x}_n + 1/\mathbf{u}_n.$$

Define $G := \{1\}$ to be the goal set, and let the safe set be its complement $S := X \setminus G$. Let us focus on the case when $\mathbf{x}_0 = 0$. If we would like to minimize the probability of reaching G over some finite horizon $n \in \mathbb{N}$, one of the optimal strategies is to choose $\mathbf{u}_0 = \frac{1}{n+1}$. Then $\mathbf{x}_1 = n+1$, $\mathbf{x}_2 = n$ and $\mathbf{x}_n = 2$, so that G is not reached. As a result, for any finite $n \in \mathbb{N}_0$ we have that $\mathsf{M}_*(0; S\mathsf{U}^n G) = 0$. However, regardless of the chosen control action $\mathbf{u}_0$, the set G is reached by the path of the process in at most $\frac{1}{\mathbf{u}_0}$ steps. Thus,

$$\mathsf{M}_*(\cdot; S\mathsf{U}G) = 1 \neq \lim_{n\to\infty} \mathsf{M}_*(0; S\mathsf{U}^n G).$$

So far we have developed DP over the value functions for the constrained reachability problem. The main tool we have used is a TC reformulation of the original performance criterion, which makes it possible to apply the rich theory that has been developed for the TC problem. Following similar lines as in the proofs of the theorems above, one can reformulate for the constrained reachability problem almost any result developed for the TC criterion. While in this paper we do not have a focus on the existence of optimal strategies, one can easily tailor a number of results from [BS78], as we overview next. Recall that Assumption (P) in [BS78, Chapter 9] corresponds to the minimization problem, whereas Assumption (N) corresponds to the maximization one.

(P) [BS78, Proposition 9.12] and its corollary provide necessary and sufficient conditions for the optimality of stationary policies, together with results to compute such policies. Moreover, [BS78, Propositions 9.17, 9.18] and their corollaries provide various sufficient conditions for the existence of optimal stationary policies, for their Borel measurability, and for the equality in (3.18).

(N) [BS78, Proposition 9.13] gives necessary and sufficient conditions for the optimality of stationary policies. However, it does not give a way to construct a policy (such as the one available for (P)). This is almost the only result on the optimality of policies under Assumption (N).

3.2. Reachability problem: computation

The TC formulation of the constrained reachability problem not only leads to results for the characterization of its solution [DAT13], but also connects to computational methods [DPR12]: both aspects are worth further investigation in future work. Alternatively, numerical methods with precise bounds on the error can be developed directly for the constrained reachability problem as in [TMKA13, Section 4]. The latter methods are based on a partitioning of the state and action spaces X and U in order to approximate the original `lcdt`-MP by a finite one. As we discussed, finite-space models are prone to be automatically verified by means of a model checker of choice. Provided certain kinds of continuity assumptions on the

kernel T, such methods assure that a bounded-horizon value function can be found with any given precision if the partition is fine enough. In the present context we are interested in extending these results to the unbounded time horizon case.

Let us recall the classical theory for the DC performance criterion. If its discounting factor satisfies $\gamma < 1$, one falls into the setting of discounted problems for which the corresponding DP operator is contractive on some function space. Such a property has nice consequences: the unbounded-horizon value function is the unique fixpoint of this operator, and it can also be efficiently approximated by means of the bounded-horizon value functions, as it follows from the contraction mapping theorem.[17] This approach is clearly interesting to us because of the computational techniques developed for the bounded time horizon case. Unfortunately, the DC formulation of the constrained reachability problem (3.10) has a discounting factor $\gamma = 1$, so the contractivity of the DP operators R^* and R_* cannot be established using classical techniques. Due to this reason, we come up with new sufficient conditions for the DP operators associated to the constrained reachability problem to be contractive: the approach is based on the following result, which is similar to that in [HL89, Proposition A.2].

Lemma 3. *Let A be any set, and let $(\mathscr{F}, \rho)$ be a metric space where $\mathscr{F}$ is any class of bounded functions $f : A \to \mathbb{R}$ and ρ is a* sup*-norm. Consider an arbitrary operator $\mathfrak{G} : \mathscr{F} \to \mathscr{F}$ that satisfies the following two properties:*

1. *if $f, g \in \mathscr{F}$ such that $f \leq g$, then $\mathfrak{G}f \leq \mathfrak{G}g$,*
2. *there exists $\beta \in [0, 1)$ such that if $f \in \mathscr{F}$ and $c \geq 0$ then $\mathfrak{G}(f + c) \leq \mathfrak{G}f + \beta c$.*

Then $\mathfrak{G}$ is a contraction on $\mathscr{F}$ with a modulus β.

Proof. Let $f, g \in \mathscr{F}$ be arbitrary, then $f \leq g + \rho(f, g)$ and thus

$$\mathfrak{G}f \leq \mathfrak{G}g + \beta\rho(f, g) \implies \mathfrak{G}f - \mathfrak{G}g \leq \beta\rho(f, g).$$

By a symmetric argument, we obtain that

$$\mathfrak{G}f - \mathfrak{G}g \leq \beta\rho(f, g) \implies |\mathfrak{G}f - \mathfrak{G}g| \leq \beta\rho(f, g) \implies \rho(\mathfrak{G}f, \mathfrak{G}g) \leq \beta\rho(f, g),$$

so that $\mathfrak{G}$ is a contraction with a modulus β. □

The DP operators for the constrained reachability problem are rarely contractive over the whole state space X, so it is worth restricting attention to the safe set S exclusively. This also emphasizes the leading role of the set S in the solution of the problem (in contrast to the goal set G), as we discussed before: we have already mentioned that the solution of the constrained reachability problem is trivial outside

[17] The contraction mapping theorem is alternatively known as Banach's Fixed Point Theorem [HL89, Proposition A.1].

of the safe set (3.2), so we can work with the restriction of value functions to the set S. Consider the "truncated" transition operator:

$$ {}_S\mathsf{T}^u f(x) := \int_S f(x')\mathsf{T}(\mathrm{d}x'|x,u), $$

which clearly maps the space $\mathscr{U}(S)$ into itself. Furthermore, let us define

$$ {}_S\mathsf{T}^* f(x) := \sup_{u\in\mathbb{K}_x} {}_S\mathsf{T}^u f(x), \qquad {}_S\mathsf{T}_* f(x) := \inf_{u\in\mathbb{K}_x} {}_S\mathsf{T}^u f(x). $$

Note that the operator ${}_S\mathsf{T}^*$ (${}_S\mathsf{T}_*$) maps the space $\mathrm{b}\mathscr{A}^*(S)$ ($\mathrm{b}\mathscr{A}_*(S)$) into itself. Moreover, for $f \in \mathrm{b}\mathscr{A}_*(X)$ it holds that $f|_S \in \mathrm{b}\mathscr{A}_*(S)$, which follows immediately from the definition of lower-semianalytic functions and Borel measurability of S; clearly, the same applies to the restrictions of functions in $\mathrm{b}\mathscr{A}^*(X)$. In particular, if we define

$$ W_n(x) := \mathsf{M}^*(x; S\mathsf{U}^n G)|_S, \qquad w_n(x) := \mathsf{M}_*(x; S\mathsf{U}^n G)|_S $$

for any $x \in X$ and $n \in \mathbb{N}_0$, then for these functions it holds that $W_n \in \mathrm{b}\mathscr{A}^*(S)$ and $w_n \in \mathrm{b}\mathscr{A}_*(S)$. Thus, we can rewrite the DP over the safe set S as follows:

$$ W_{n+1} = {}_S\mathsf{R}^*[W_n], \qquad w_{n+1} = {}_S\mathsf{R}_*[w_n] $$

for any $n \in \bar{\mathbb{N}}_0$, where $W_0 = w_0 = 0$, and the truncated DP operators are given by

$$ {}_S\mathsf{R}^* f(x) := \sup_{u\in\mathbb{K}_x} [\mathsf{T}(G|x,u) + {}_S\mathsf{T}^u f(x)], \qquad f \in \mathrm{b}\mathscr{A}^*(S), $$
$$ {}_S\mathsf{R}_* f(x) := \inf_{u\in\mathbb{K}_x} [\mathsf{T}(G|x,u) + {}_S\mathsf{T}^u f(x)], \qquad f \in \mathrm{b}\mathscr{A}_*(S). $$

Clearly, these operators map their domains into themselves, so that they can be applied recursively. Note also that in case $G = \emptyset$, we have ${}_S\mathsf{R}^u = {}_S\mathsf{T}^u$.

In order to formulate the main result on the contractivity of the DP operators, we are only left to introduce a very important special case of constrained reachability: safety [APLS08]. This can be characterized by the LTL formula $\Box^n S$ and thus

$$ \mathsf{M}^\pi(x; \Box^n S) = 1 - \mathsf{M}^\pi(x; S\mathsf{U}^n S^c) $$

for all $x \in X$ and any $n \in \bar{\mathbb{N}}_0$. We are interested in the restriction of the safety problem to the safe set S itself, the main focus being the characterization of contractivity.[18] We further denote

$$ V_n(x) := \mathsf{M}^*(x; \Box^n S)|_S, \qquad v_n(x) := \mathsf{M}_*(x; \Box^n S)|_S. $$

The DP for safety over S is hence given by

$$ V_{n+1} = {}_S\mathsf{T}^*[V_n], \qquad v_{n+1} = {}_S\mathsf{T}_*[v_n], \quad n \in \bar{\mathbb{N}}_0. $$

[18] Clearly, $V_n^\pi(x; S) = 0$ for any $x \in S^c$, so the safety problem is trivial outside the safe set.

with $V_0 = \nu_0 = 1$. Clearly, we have that $0 \le V_n \le 1$ for all $n \in \bar{\mathbb{N}}_0$. Let us define

$$\beta_n(S) := \sup_{x \in S} V_n(x) = \sup_{x \in X} \mathsf{M}^*(x; \Box^n S) \in [0,1],$$

$$m(S) := \inf\{n \in \mathbb{N}_0 : \beta_n(S) < 1\} \in \bar{\mathbb{N}}_0,$$

and note that both β_n and m are monotonic functions of S with respect to set inclusion. We are now ready to provide sufficient conditions for contractivity.

Theorem 4. *If $m := m(S) < \infty$, then operators $({}_S\mathsf{R}^*)^m$ and $({}_S\mathsf{R}_*)^m$ are contractions with modulus $\beta := \beta_m(S)$ on the spaces $\mathrm{b}\mathscr{A}^*(S)$ and $\mathrm{b}\mathscr{A}_*(S)$ respectively. In particular, each of them has a unique fixpoint, and for any $n \in \mathbb{N}_0$ the following inequalities hold true:*

$$\rho(W_\infty, W_{mn}) \le \beta^n, \quad \rho(w_\infty, w_{mn}) \le \beta^n. \tag{3.20}$$

Finally, as a special case $({}_S\mathsf{T}^*)^m$ and $({}_S\mathsf{T}_*)^m$ are contractions and $V_\infty = \nu_\infty = 0$.

Proof. We are going to apply Lemma 3 in order to establish the contractivity property. Let us consider the case of ${}_S\mathsf{R}^*$ first, so in Lemma 3 we put $\mathscr{F} = \mathrm{b}\mathscr{A}^*(S)$. The condition (1) of the lemma is obviously satisfied for ${}_S\mathsf{R}^*$ and hence for $({}_S\mathsf{R}^*)^n$ regardless of $n \in \mathbb{N}_0$. Furthermore, for any two functions $f, g \in \mathrm{b}\mathscr{A}^*(S)$ we have that

$$\begin{aligned}
{}_S\mathsf{R}^*(f(x) + g(x)) &= \sup_{u \in \mathbb{K}_x} [\mathsf{T}(G|x,u) + {}_S\mathsf{T}^u f(x) + {}_S\mathsf{T}^u g(x)] \\
&\le \sup_{u \in \mathbb{K}_x} [\mathsf{T}(G|x,u) + {}_S\mathsf{T}^u f(x)] + \sup_{u \in \mathbb{K}_x} {}_S\mathsf{T}^u g(x) \\
&= {}_S\mathsf{R}^* f(x) + {}_S\mathsf{T}^* g(x).
\end{aligned}$$

As a result, for any $f \in \mathrm{b}\mathscr{A}^*(S)$ and any $c \ge 0$ it holds that

$${}_S\mathsf{R}^*(f + c) \le {}_S\mathsf{R}^* f + c \cdot V_1,$$

and further by induction for any $n \in \mathbb{N}$

$$({}_S\mathsf{R}^*)^n(f + c) \le ({}_S\mathsf{R}^*)^n f + c \cdot V_n.$$

In particular, for the case $n = m$ we obtain the following:

$$({}_S\mathsf{R}^*)^m(f + c) \le ({}_S\mathsf{R}^*)^m f + c \cdot V_m \le ({}_S\mathsf{R}^*)^n f + c \cdot \beta.$$

Hence, $({}_S\mathsf{R}^*)^m$ satisfies all the assumptions of Lemma 3 and thus is a contraction on $\mathrm{b}\mathscr{A}^*(S)$. The contractivity of $({}_S\mathsf{R}_*)^m$ can be shown by a similar argument, with the only difference being the inequality

$${}_S\mathsf{R}_*(f + g) \le {}_S\mathsf{R}_* f + {}_S\mathsf{T}_* g,$$

rather than the one with ${}_S\mathsf{T}^* g$, and with conditions on contractivity that are stated in terms of functions ν_n rather than V_n.

After the contractivity of the operators is established, the uniqueness of the solutions of fixpoint equations and the bounds in (3.20) follow immediately from the contraction mapping theorem [HL89, Proposition A.1]. Finally, the statement for operators ${}_S\mathsf{T}^*$ and ${}_S\mathsf{T}_*$ follows directly if one considers the special case $G = \emptyset$ when the two coincide. □

Theorem 4 shows that in the case of contractive operators the unbounded-horizon value function can be approximated by n-bounded horizon ones with any precision level that is a function of the modulus β and on the horizon n. However, there are some questions left: what are the cases in which the contractivity conditions are violated, and what would be a solution for such cases? Let us first address the former question. For example, whenever the conditions of Theorem 4 are met, the equality holds in (3.18). As a result, Example 1 does not admit contractive operators since the equality does not hold there. Some other important examples can be given using the notion of absorbing set.

Definition 7 (Absorbing set)**.** *The set $A \in \mathscr{B}(X)$ is called* strongly absorbing *if $\mathsf{T}(A|x,u) = 1$ for all $x \in A$ and $u \in \mathbb{K}_x$. The set $A \in \mathscr{B}(X)$ is called* weakly absorbing *if there exists a randomized selector $\mu \in \mathscr{U}(U|X)$ such that for all $x \in A$ it holds that $\mu(\mathbb{K}_x|x) = 1$ and that*

$$\int_{\mathbb{K}_x} \mathsf{T}(A|x,u)\mu(\mathrm{d}u|x) = 1. \tag{3.21}$$

We say that the set $A \in \mathscr{B}(X)$ is simple *if it does not have non-empty weakly absorbing subsets.*

The following notation is extensively used below: for any $A \in \mathscr{B}(X)$ we define

$$\mathbb{K}^A := \{(x,u) \in \mathbb{K} : \mathsf{T}(A|x,u) = 1\},$$

and analogously $\mathbb{K}^A_x, x \in A$. Note in particular that if the sets $A, B \in \mathscr{B}(X)$ are such that $B \subseteq A$, then $\mathsf{T}(B|x,u) = 1$ for some $(x,u) \in \mathbb{K}$ implies that $\mathsf{T}(A|x,u) = 1$, and as a result we obtain that $\mathbb{K}^B \subseteq \mathbb{K}^A$. The next theorem establishes some important results on the connection between weakly and strongly absorbing sets, and on their structure.

Proposition 5. *Let $A \in \mathscr{B}(X)$. It holds that*

i. if A is strongly absorbing, then it is weakly absorbing,

ii. if A is weakly absorbing, then the randomized selector μ in (3.21) can be equivalently replaced by a deterministic selector $f \in \mathscr{U}(X)/\mathscr{B}(U)$.

Proof. To prove i. we use the fact that $\mathsf{T}(A|x,\mathsf{k}(x)) = 1$ for any $x \in A$ (recall that $\mathsf{k} : X \to U$). Hence, the kernel μ as per (3.21) can be chosen to be a deterministic selector, as $\mu = \delta_{\mathsf{k}}$.

With focus on ii. let us fix some arbitrary $x \in A$ and show that there exists $u \in \mathbb{K}_x$ such that $\mathsf{T}(A|x,u) = 1$. Note that if $u \notin \mathbb{K}^A_x$ (cf. notations above), then $1 - T(A|x,u) > 0$, where it is crucial that the inequality is strict. To reach a contradiction we further suppose that for a μ as in (3.21) it holds that $\mu(\mathbb{K}_x \setminus \mathbb{K}^A_x|x) > 0$. Then:

$$0 = \int_{\mathbb{K}_x} (1 - \mathsf{T}(A|x,u))\mu(\mathrm{d}u|x) \geq \int_{\mathbb{K}_x \setminus \mathbb{K}^A_x} (1 - \mathsf{T}(A|x,u))\mu(\mathrm{d}u|x) > 0,$$

which obviously cannot be true. As a result, we obtain that $\mu(\mathbb{K}_x^A|x) = 1$ and in particular $\mathbb{K}_x^A \neq \emptyset$ for any $x \in A$. Hence, it holds that $\mathsf{T}^* 1_A(x) = \sup_{u \in \mathbb{K}_x} \mathsf{T}(A|x,u) = 1$. The existence of a universally measurable selector u from $\mathbb{K}_x^A$ thus follows from [BS78, Proposition 7.50 (b)] and the fact that $\mathsf{T}(A|\cdot) \in \mathrm{b}\mathscr{B}(\mathbb{K}) \subseteq \mathrm{b}\mathscr{A}^*(\mathbb{K})$. □

Part $i.$ of Proposition 5 justifies the use of the adjectives "weak" and "strong" in Definition 7. Furthermore, in the uncontrolled case (where trivially $U = \{u\}$), the notion of weak and strong sets coincide with that of an absorbing set [MT93]. Intuitively, a strongly absorbing set remains absorbing under any possible control action, whereas for a weakly absorbing set there has to exist a control policy that makes this set absorbing. Moreover, thanks to part $ii.$ of Proposition 5, it is sufficient to consider non-randomized controls in order to establish weak absorbance.

As promised, absorbing sets can be used to provide examples when the contractivity of truncated DP operators is violated, and in particular when the fixpoint equations do not have unique solutions. Note that in the case of unconstrained reachability $G = S^c$, it holds that the operators ${}_S\mathsf{R}^*$ and ${}_S\mathsf{R}_*$ always admit the trivial fixpoint 1. However, if S is not simple (that is, if it admits absorbing subsets), then the optimal value functions are different than 1. For example, if a trajectory starts in an absorbing subset of S then it never reaches the goal set. More precisely:

Proposition 6. *If a set S has a non-empty strongly (weakly) absorbing subset $A \subseteq S$, then $\mathsf{M}^*(x; S \mathsf{U} S^c) = 0$ ($\mathsf{M}_*(\cdot; S \mathsf{U} S^c) = 0$) for all $x \in A$. In particular, $W_\infty(x) = 0$ ($w_\infty(x) = 0$) for all $x \in A$, and $({}_S\mathsf{R}^*)^n$ ($({}_S\mathsf{R}_*)^n$) is not a contraction for any $n \in \mathbb{N}_0$.*

Proof. Let A be a strongly absorbing set and fix a point $x \in A$. Then $\mathsf{P}_x^\pi(\mathbf{x}_n \in A) = 1$ for all $n \in \mathbb{N}_0$ regardless of the policy $\pi \in \Pi$. As a result, $\mathsf{P}_x^\pi(\mathbf{x}_n \in S^c) = 0$ for all $n \in \mathbb{N}_0$, so

$$\mathsf{M}^\pi(x; S \mathsf{U} S^c) \leq \sum_{n=0}^{\infty} \mathsf{P}_x^\pi(\mathbf{x}_n \in S^c) = 0$$

for any policy $\pi \in \Pi$. Thus, we obtain that $\mathsf{M}^*(x; S \mathsf{U} S^c) = 0$ for any $x \in A$. Clearly, it follows immediately that $W_\infty(x) = 0$ for all $x \in A$. Suppose now that $({}_S\mathsf{R}^*)^n$ is contractive for some n. In such a case the solution of the fixpoint equation would be unique and hence it would imply that $W_\infty \equiv 1$, which is clearly not the case.

Let now A be a weakly absorbing set and consider a stationary policy $\pi \in \Pi_S$ with

$$\pi_0(x) := 1_A(x) \cdot \mu(x) + 1_{A^c}(x) \cdot \delta_{\mathsf{k}}(x),$$

with μ as in (3.21). The policy π uses the choice of the action suggested by μ whenever $x \in A$, and chooses some auxiliary action $\mathsf{k}(x)$ otherwise. From the definition of μ it follows that $\mathsf{P}_x^\pi(\mathbf{x}_n \in A) = 1$ and hence $\mathsf{P}_x^\pi(\mathbf{x}_n \in S^c) = 0$ for all $x \in A$, so

$$\mathsf{M}_*(\cdot; S \mathsf{U} S^c) \leq \mathsf{M}^\pi(x; S \mathsf{U} S^c) \leq \sum_{n=0}^{\infty} \mathsf{P}_x^\pi(\mathbf{x}_n \in S^c) = 0.$$

As for ${}_S\mathsf{R}_*$, one can now show that $({}_S\mathsf{R}_*)^n$ is not a contraction for any $n \in \mathbb{N}_0$. □

In general the presence of absorbing sets is not the only reason that may violate contractivity. For example, it is easy to see that the set S in Example 1 does not have weakly absorbing subsets, and still the contractivity does not hold. However, the following assumption allows to characterize precisely the relationship between absorbing sets and contractivity.

Assumption 1. *The* cdt*-MP $\mathfrak{D}$ is continuous and the set S is compact.*

We are going to show that, under Assumption 1, the case $m(S) < \infty$ precisely coincides with the case when S does not admit weakly absorbing sets. In order to prove this fact some preparation is required: let us define for all $n \in \mathbb{N}_0$ the sets

$$S_n := \{\mathsf{M}^*(\cdot, \Box^n S) = 1\} = \{x \in X : \mathsf{M}^*(x, \Box^n S) = 1\}.$$

Note that for any $x \in S$ and $\pi \in \Pi$ the sequence $(\mathsf{M}^\pi(x; \Box^n S))_{n\in\mathbb{N}_0}$ is non-increasing, as is the sequence $(\mathsf{M}^*(x; \Box^n S))_{n\in\mathbb{N}_0}$. As a result, we obtain that the sequence of sets $(S_n)_{n\in\mathbb{N}_0}$ is non-increasing as well: $S_{n+1} \subseteq S_n$ for all $n \in \mathbb{N}_0$. Let us further denote by $S_\infty := \bigcap_{n=0}^\infty S_n$ the limit of this sequence. We introduce the following auxiliary lemmas.

Lemma 4. *The set S_∞ is such that $\{\mathsf{M}^*(\cdot; \Box S) = 1\} \subseteq S_\infty$. In particular, if $S' \subseteq S$ is a weakly absorbing subset of S, then $S' \subseteq S_\infty$.*

Proof. Let us fix any x such that $\mathsf{M}^*(x; \Box S) = 1$. By Proposition 4 we have that

$$\lim_{n\to\infty} \mathsf{M}^*(x; \Box^n S) \geq \mathsf{M}^*(x; \Box S) = 1.$$

Since $(\mathsf{M}^*(x; \Box^n S))_{n\in\mathbb{N}_0}$ is a non-increasing sequence, it follows that $\mathsf{M}^*(x; \Box^n S) = 1$ and hence $x \in S_n$ for all $n \in \mathbb{N}_0$. As a result, $x \in S_\infty$ and thus $\{\mathsf{M}^*(\cdot; \Box S) = 1\} \subseteq S_\infty$. Now, if $S' \subseteq S$ is weakly absorbing, then $S' \subseteq \{\mathsf{M}^*(\cdot; \Box S) = 1\}$ by Proposition 6. □

Lemma 5. *Under Assumption 1 it holds that $\mathsf{M}^*(\cdot; \Box^n S) \in \mathrm{b}\mathscr{C}^*(X)$ for all $n \in \mathbb{N}_0$.*

Proof. If $n = 0$ then $\mathsf{M}^*(x; \Box^0 S) = 1_S(\cdot) \in \mathrm{b}\mathscr{C}^*(X)$ since S is a closed set being a compact subset of a metrizable space. Also, if $\mathsf{M}^*(\cdot; \Box^n S) \in \mathrm{b}\mathscr{C}^*(X)$, then by continuity of the kernel T we have that $\mathsf{T}\mathsf{M}^*(\cdot; \Box^n S) \in \mathrm{b}\mathscr{C}^*(\mathbb{K})$ and $\mathsf{T}^*\mathsf{M}^*(\cdot; \Box^n S) \in \mathrm{b}\mathscr{C}^*(X)$ as it follows from [BS78, Proposition 7.31] and [BS78, Proposition 7.33] respectively. Finally, $\mathsf{M}^*(\cdot; \Box^{n+1} S) = 1_S(\cdot) \cdot \mathsf{T}^*\mathsf{M}^*(\cdot; \Box^n S) \in \mathrm{b}\mathscr{C}^*(X)$ by Lemma 11 in the Appendix. □

Lemma 6. *Under Assumption 1, sets S_n and $\mathbb{K}_x^{S_n}$ are compact for all $x \in X$, $n \in \mathbb{N}_0$.*

Proof. Since $S_n = \{\mathsf{M}^*(\cdot; \Box^n S) \geq 1\}$ and $\mathsf{M}^*(\cdot; \Box^n S)$ is an upper semi-continuous function by Lemma 5, we obtain that S_n is a closed set. It is also compact as a closed subset of a compact set S. Furthermore, it holds that $\mathsf{T}1_{S_n} \in \mathrm{b}\mathscr{C}^*(\mathbb{K})$ since the set S_n is closed. Hence, $\mathbb{K}^{S_n} = \{\mathsf{T}1_{S_n} \geq 1\}$ is a closed subset of $\mathbb{K}$, which implies that $\mathbb{K}^{S_n}(x)$ is a closed subset of U for any $x \in X$, and is compact since U is compact. □

Lemma 7. *Under Assumption 1, $S_{n+1} = \{x \in S : \mathbb{K}^{S_n}(x) \neq \emptyset\}$ for any $n \in \mathbb{N}_0$, that is*

$$S_{n+1} = \{x \in S : \exists u \in \mathbb{K}_x \text{ s.t. } \mathsf{T}(S_n|x,u) = 1\}. \tag{3.22}$$

Moreover, S_∞ is weakly absorbing and satisfies the formula $S_\infty = \{\mathsf{M}(\cdot;\Box S) = 1\}$.

Proof. Let us first prove (3.22) for $n < \infty$. We first show that if $\mathbb{K}^{S_n}(x) \neq \emptyset$ for some $x \in S$, then $x \in S_{n+1}$. Indeed, let u' be an arbitrary element of $\mathbb{K}^{S_n}(x)$. We have:

$$\mathsf{M}^*(x;\Box^{n+1}S) = \sup_{u\in\mathbb{K}_x} \int_X \mathsf{M}^*(x';\Box^n S)\mathsf{T}(\mathrm{d}x'|x,u) \geq \int_{S_n} \mathsf{M}^*(x';\Box^n S)\mathsf{T}(\mathrm{d}x'|x,u') = 1,$$

so that $\mathbb{K}^{S_n}(x) \neq \emptyset$ for $x \in S$ implies $x \in S_{n+1}$. Showing the converse implication is more technical. Let us pick $x \in S$ arbitrarily. Since $\mathsf{T}\mathsf{M}^*(\cdot;\Box^n S) \in \mathrm{b}\mathscr{C}^*(\mathbb{K})$ as a function of x and u, it follows that $\mathsf{T}\mathsf{M}^*(\cdot;\Box^n S) \in \mathrm{b}\mathscr{C}^*(\mathbb{K}_x)$ as a function of u, and thus the maximum in

$$\mathsf{M}^*(x;\Box^{n+1}S) = \sup_{u\in\mathbb{K}_x} \mathsf{T}\mathsf{M}^*(x,u;\Box^n S)$$

is attained at some $u'' \in \mathbb{K}_x$ as the latter set is compact. As a result,

$$\int_S 1\, \mathsf{T}(\mathrm{d}x'|x,u'') = \mathsf{T}(S|x,u'') \leq 1 = \mathsf{M}^*(x;\Box^{n+1}S) = \int_S \mathsf{M}^*(x';\Box^n S)\mathsf{T}(\mathrm{d}x'|x,u''),$$

and subtracting the right-hand side from the left-hand side yields

$$\int_S \left[1 - \mathsf{M}^*(x';\Box^n S)\right]\mathsf{T}(\mathrm{d}x'|x,u'') \leq 0. \tag{3.23}$$

Note that the integrand in (3.23) is non-negative, so

$$\mathsf{T}(\{1 - \mathsf{M}^*(\cdot;\Box^n S) = 0\}|x,u'') = \mathsf{T}(S_n|x,u'') = 1,$$

since otherwise the integral would be strictly positive. Thus, (3.22) is proved.

With focus on S_∞, the case when S_∞ is empty is trivial, so let us assume $S_\infty \neq \emptyset$. For any $x \in S_\infty$ it follows that $x \in S_n \neq \emptyset$ for all $n \in \mathbb{N}_0$ and hence $\mathbb{K}^{S_n}(x) \neq \emptyset$ for all $n \in \mathbb{N}_0$. Indeed, in case $\mathbb{K}^{S_n}(x) = \emptyset$ we would obtain that $x \notin S_{n+1}$ thanks to (3.22) which contradicts with the fact that $x \in S_\infty$. Then $\mathbb{K}^\infty(x) := \bigcap_{n=0}^\infty \mathbb{K}^{S_n}(x) \neq \emptyset$ since $(\mathbb{K}^{S_n}(x))_{n\in N_0}$ is a non-increasing sequence of non-empty compact sets. For any $u' \in \mathbb{K}^\infty(x)$ and any $n \in \mathbb{N}_0$ it holds that $\mathsf{T}(S_n|x,u') = 1$ so that

$$\mathsf{T}(S_\infty|x,u') = \mathsf{T}\left(\bigcap_{n=0}^\infty S_n|x,u'\right) = \lim_{n\to\infty} \mathsf{T}(S_n|x,u') = 1$$

and so $x \in S_\infty$ implies that $\mathbb{K}^{S_\infty}(x) \neq \emptyset$. As a result, we obtain that for all $x \in S_\infty$

$$\mathsf{T}^* 1_{S_\infty}(x) = \sup_{u\in\mathbb{K}_x} \mathsf{T}(S_\infty|x,u) = 1.$$

The set S_∞ is compact as an intersection of compact sets, so that $\mathsf{T}(S_\infty|\cdot) \in \mathrm{b}\mathscr{C}^*(\mathbb{K})$, and thus by [BS78, Proposition 7.33] there exists a selector $f : \mathscr{B}(X)/\mathscr{B}(U)$ such that $f(x) \in \mathbb{K}_x$ for all $x \in X$ and $\mathsf{T}(S_\infty|x, f(x)) = 1$ for all $x \in S_\infty$. Hence, S_∞ is a weakly absorbing subset and thus $S_\infty \subseteq \{\mathsf{M}^*(\cdot;\Box S) = 1\}$. Combining the latter statement with the result of Lemma 4, we obtain that $S_\infty = \{\mathsf{M}^*(\cdot;\Box S) = 1\}$. □

Lemma 7 shows that when the `cdt`-MP $\mathfrak{D}$ is continuous, any compact set S admits a largest weakly absorbing subset S_∞ (which clearly may be empty). We are now able to show that this is equivalent to the contractivity condition.

Theorem 5. *Under Assumption 1, the following statements are equivalent:*

i. it holds that $m(S) < \infty$ (contractivity);

ii. the operator ${}_S\mathsf{T}^*$ has a unique fixpoint (uniqueness);

iii. it holds that $\mathsf{M}^*(\cdot;\Box S) = 0$ (triviality);

iv. it holds that $S_\infty = \emptyset$ (simplicity).

Proof. The fact that $i. \implies ii.$ has been proven in Theorem 4. Also, ${}_S\mathsf{T}^* f = f$ always has a solution $f = 0$, so the uniqueness of a fixpoint of ${}_S\mathsf{T}^*$ implies $V_\infty = 0$ and thus $ii. \implies iii.$ If $\mathsf{M}^*(\cdot;\Box S) = 0$, then by Lemma 7 we have $S_\infty = \{\mathsf{M}^*(\cdot;\Box S) = 1\} = \emptyset$, so $iii. \implies iv.$ Finally, if $m(S) = \infty$ then $\sup_{x\in X} \mathsf{M}^*(x;\Box^n S) = 1$ for all $n \in N_0$. By Lemma 5 each of the functions $\mathsf{M}^*(\cdot;\Box^n S)$ is u.s.c. and thus it attains its maximum over a compact set S, so that $m(S) = \infty$ implies $S_n \neq \emptyset$ for all $n \in \mathbb{N}_0$. Moreover, from Lemma 6 it follows that each S_n is compact and hence $(S_n)_{n\in\mathbb{N}_0}$ is a non-increasing sequence of non-empty compact sets. As a result, the latter sequence has a non-empty intersection S_∞ and hence $S_\infty = \emptyset$ necessarily implies $m(S) < \infty$, hence $iv. \implies i.$ □

We have obtained a precise characterization of the contractivity condition $m(S) < \infty$ in terms of presence or absence of weakly absorbing subsets of the safe set. In particular, if both Assumption 1 and the condition $S_\infty = \emptyset$ are satisfied, then regardless of the set G we are able to approximate $\mathsf{M}^*(x;S\mathsf{U}G)$ and $\mathsf{M}_*(x;S\mathsf{U}G)$ by their bounded horizon counterparts. Moreover, Theorem 5 also justifies the following intuitive statement: if one wants to keep the path of the process inside a set with some non-zero probability, there has to be an "attractor" within such set, which in our case appears to be the largest weak absorbing subset of S, that is S_∞. If such attractor is absent, no matter what control policy is chosen, the path will leave the desired set almost surely. The "if and only if" nature of Theorem 5 also implies that for the maximal safety problem such condition is necessary. However, it still may be the case that $S_\infty \neq \emptyset$ but ${}_S\mathsf{R}_*$ is a contraction. Although such cases are interesting to study, this goes beyond the scope of the current paper: we are now interested in techniques that allow us to reduce the unbounded horizon problem to the bounded horizon one in the situation where $S_\infty \neq \emptyset$. These results are particularly powerful under the following assumption.

Assumption 2. *Stationary policies are sufficient for the solution of the constrained reachability problem on the unbounded time horizon, that is for any $x \in X$:*

$$\mathsf{M}^*(x; S \mathsf{U} G) = \sup_{\pi \in \Pi_S} \mathsf{M}^\pi(x; S \mathsf{U} G), \qquad \mathsf{M}_*(x; S \mathsf{U} G) = \inf_{\pi \in \Pi_S} \mathsf{M}^\pi(x; S \mathsf{U} G).$$

Before we provide the main result, the following technical lemma is needed.

Lemma 8. *Let Assumption 2 hold true, and let $C \in \mathscr{B}(X)$ be any subset of S. Then $\forall x \in X$*

$$|\mathsf{M}^*(x; S \mathsf{U} G) - \mathsf{M}^*(x; (S \setminus C) \mathsf{U} G)| \leq \chi^*(C) := \sup_{\pi \in \Pi_S} \sup_{y \in C} \mathsf{M}^\pi(y; S \mathsf{U} G),$$
$$|\mathsf{M}_*(x; S \mathsf{U} G) - \mathsf{M}_*(x; (S \setminus C) \mathsf{U} G)| \leq \chi_*(C) := \inf_{\pi \in \Pi_S} \sup_{y \in C} \mathsf{M}^\pi(y; S \mathsf{U} G).$$

Proof. For any $\pi \in \Pi_S$ let us denote $\chi^\pi(C) := \sup_{y \in C} \mathsf{M}^\pi(y; S \mathsf{U} G)$. Let us fix an arbitrary policy $\pi \in \Pi_S$ and an arbitrary state $x \in X$. Clearly, $S \setminus C \subseteq S$ further implies that $\mathsf{M}^\pi(x; S \mathsf{U} G) \geq \mathsf{M}^\pi(x; (S \setminus C) \mathsf{U} G)$. On the other hand, obviously

$$\mathsf{M}^\pi(\cdot; S \mathsf{U} G) - \mathsf{M}^\pi(\cdot; (S \setminus C) \mathsf{U} G) \leq \chi^\pi(C)$$

which together with Lemma 10 immediately yields the desired result. □

Let us discuss how Lemma 8 can be useful. Suppose that Assumption 1 holds true and that for the original problem we have that $S_\infty \neq \emptyset$, so that $m(S) = \infty$, and hence we cannot apply Theorem 4 to compute the optimal value functions. If we find a set $C \supseteq S_\infty$ such that $m(S \setminus C) < \infty$, then we can solve the unconstrained problem with truncated safe set $S \setminus C$. Also, since C contains S_∞ we can expect that $\chi^*(C)$ and $\chi_*(C)$ are close enough to zero, which would make the bounds in Lemma 8 useful. To further elaborate this idea we need the notion of a *locally excessive* function.

Definition 8 (Locally excessive function)**.** *A non-negative function $g \in \mathrm{b}\mathscr{B}(X)$ is called* locally μ-excessive *for a randomized selector $\mu \in \mathscr{U}(U|X)$, if for any $x \in \{g \leq 1\}$ it holds that $\mathsf{T}^\mu g(x) \leq g(x)$. If in addition $A_\infty \subseteq \{g = 0\}$ and $\{g \leq 1\} \subseteq A$ for some $A \in \mathscr{B}(X)$, and $\{g < \varepsilon\}$ is an open set for all $\varepsilon > 0$, we say that g is locally μ-excessive on A.*

A non-negative function $g \in \mathrm{b}\mathscr{B}(X)$ is called locally uniformly excessive if for any $x \in \{g \leq 1\}$ and $u \in \mathbb{K}_x$ it holds that $(\mathsf{T} g)(x, u) \leq g(x)$. If in addition $A_\infty \subseteq \{g = 0\}$ and $\{g \leq 1\} \subseteq A$ for some $A \in \mathscr{B}(X)$, and $\{g < \varepsilon\}$ is an open set for all $\varepsilon > 0$, we say that g is locally uniformly excessive on A.

The next results allows characterising, via simple sets, the solution of reachability problems.

Theorem 6. *Let Assumptions 1 and 2 hold true. Suppose that g^* is locally uniformly excessive on S, and that g_* is locally π'_0-excessive for some $\pi' \in \Pi_S$. For any $\varepsilon \in (0, 1]$ it holds that the following inequalities are valid:*

$$\chi^*(\{g^* < \varepsilon\}) \leq \varepsilon \qquad \chi_*(\{g_* < \varepsilon\}) \leq \varepsilon$$

and that sets $S \setminus \{g^* < \varepsilon\}$, $S \setminus \{g_* < \varepsilon\}$ are simple.

Proof. We start with the case of the maximization. For any policy $\pi \in \Pi_S$ we have that

$$\mathsf{M}^{\pi}(x; X \mathsf{U} \{g^* > 1\}) \leq g^*(x)$$

whenever $x \in \{g^* \leq 1\}$, as it follows from [TA14, Lemma 3]. Furthermore, since $\{g^* \leq 1\} \subseteq S$ and $G \subseteq S^c$, it holds that $G \subseteq \{g^* > 1\}$. As a result,

$$\mathsf{M}^{\pi}(\cdot; S \mathsf{U} G) \leq \mathsf{M}^{\pi}(\cdot; X \mathsf{U} \{g^* > 1\}).$$

Combining both inequalities, we obtain that

$$\sup_{y \in \{g^* < \varepsilon\}} \mathsf{M}^{\pi}(y; S \mathsf{U} G) \leq \varepsilon,$$

and thus after maximizing over all stationary policies we obtain that $\chi^*(\{g^* < \varepsilon\}) \leq \varepsilon$.

For the case of the minimization we similarly have

$$\sup_{y \in \{g_* < \varepsilon\})} \mathsf{M}^{\pi'}(y; S \mathsf{U} G) \leq \varepsilon,$$

and since $\chi_*(C) \leq \sup_{y \in C} \mathsf{M}^{\pi'}(y; S \mathsf{U} G)$ for any set $C \in \mathscr{B}(X)$, we immediately obtain that $\chi_*(\{g_* < \varepsilon\}) \leq \varepsilon$ for all $\varepsilon \leq 1$, as desired.

Finally, the simplicity of sets $S \setminus \{g^* < \varepsilon\}$ and $S \setminus \{g_* < \varepsilon\}$ follows from the fact that that they are compact simple sets. Indeed, we have compactness thanks to the fact that S is compact and sets $\{g^* < \varepsilon\}$, $\{g_* < \varepsilon\}$ are open. Moreover, the simplicity follows from the definition of functions locally excessive on S which implies that $S_\infty \subseteq \{g^* < \varepsilon\}$ and $S_\infty \subseteq \{g_* < \varepsilon\}$. □

3.3. Comments on the reachability problem

Let us mention how the DP formulation has been developed for the (un)constrained reachability problem in the `cdt`-MP setting. To our knowledge, the first work with this goal has been [APLS08], which has considered a class of models called controlled discrete-time Stochastic Hybrid Systems (`cdt`-SHS), namely a class of `cdt`-MP with a state space comprised of a collection of Borel subsets of $\mathbb{R}^n$. It has treated the unconstrained reachability property $\Diamond^n G = \mathtt{true} \mathsf{U}^n G$ and the dual safety one $\Box^n S = \neg \Diamond^n S^c$, and has proposed their characterization using a maximal cost (3.3) for the first problem, and a multiplicative cost (3.4) for the second. Within this formulations, the DP recursion has been derived for the bounded time horizon $n < \infty$, while restricting the attention to Markov policies. [SL10] has addressed a more general[19] constrained reachability problem $S \mathsf{U}^n G$ within a similar setting: `cdt`-SHS

[19] Although constrained reachability includes the unconstrained one as a special case, the latter can be used to solve the former if one just slightly modifies dynamic by making the set of unsafe sets $D = X \setminus (S \cup G)$ absorbing. Indeed, in such case $\Diamond^n G$ is equivalent to $S \mathsf{U}^n G$ since G is never reached by a trajectory that has visited D at least once [TMKA13, Proposition 1]. In particular, one immediately obtains [SL10, Theorem 8] by applying [APLS08, Theorem 1] over a modified model. Similarly, rendering the set D absorbing allows one to recast a related terminal hitting-time reach-avoid problem [SL10, Section 4] as a special case of a terminal cost problem [HLL96, Section 3].

models, Markov policies, and bounded time horizons: a new sum-multiplicative cost (3.5) has been proposed, leading to the DP scheme in [SL10, Theorem 8]. In contrast to these studies, here we have proposed a TC formulation, which has allowed dealing with non-Markovian policies, and to show that Markov policies are sufficient. In particular, one obtains [APLS08, Theorems 1, 2] and [SL10, Theorem 8] as special cases of Theorem 3. At the same time, the TC formulation has also led to simpler proofs, which mostly rely on known results for the TC performance criterion [BS78, Chapters 8,9].

The case of the unbounded time horizon problem has received some attention already in [SL10, Section 3.3] and [AAP+06, Section V]. There it was suggested to use the convergence of the bounded-horizon values to the unbounded-horizon one, which led to considering the fixpoint equations. Although we have shown in Theorem 3 that fixpoint equations are indeed valid, they can not be obtained using limiting arguments as the latter may fail as shown in Example 1. An alternative approach via a hitting time formulation (3.6) has been proposed in [CCL11], and the fixpoint equation for the maximal constrained reachability has been obtained in [CCL11, Theorem 2.10 (i)]. However, one of the assumptions of this theorem required the first hitting time of the complement of the safe set (call it τ_{S^c}) to be almost surely finite for any Markov policy. As a result, in the case of unconstrained reachability this result requires to be properly adapted. Finally, [KSL13, Theorem 2] has shown the convergence of the *maximal* bounded-horizon unconstrained reachability to the unbounded-horizon one, and has showed that the latter satisfies the fixpoint equation. In contrast to the aforementioned contributions, Theorem 3 does not pose any limitations and establishes fixpoint equations for both the maximization and the minimization problems in generality, without for example requiring any continuity assumptions that are often imposed otherwise (cf. [KSL13, Assumption 1] or [CCL11, Assumption 2.9]). In addition, Proposition 4 provides a complete characterization of the convergence of bounded-horizon problems to the unbounded-horizon ones, and is further supported by Example 1.

The approximation of the unbounded-horizon reachability problem with bounded-horizon counterparts is an extension to the controlled case of the result in [TA14]. This extension requires no additional assumptions and (weak) continuity of the kernel T is sufficient to establish important results such as Theorems 5 and 6. At the same time, in the proofs we have extensively used continuity assumption, and so the equivalence in Theorem 5 may fail to hold without such assumptions – see e.g. [TA14, Appendix]. In particular, we acknowledge that [TMKA13, Proposition 2] is not correct: although uniqueness of fixpoint indeed yields trivial constant solutions for the maximal and minimal unconstrained reachability in the general case, without continuity assumptions it may happen that the solution is trivial but yet there are multiple fixpoints. In emphasizing the role of absorbing sets, it is crucial to use the connection between $m(S)$ and the contractivity of powers of the operators ${}_S\mathbb{R}^*$ and ${}_S\mathbb{R}_*$ in Theorem 4. In particular, as a special case we obtain [KSL13, Proposition 1], which has obtained conditions for the contractivity in the special case $m(S) = 1$. The characterization of the absorbing sets, as well as finding an appropriate μ-excessive function, is an interesting and important problem. For example, there seems to be a connection between weakly absorbing sets (such as S_∞)

and maximal controlled invariant sets in non-stochastic systems [RMT13]. Another related concept is that of the maximal end component (MEC) [BK08, Section 10.6], which is used to solve both the reachability and repeated reachability problems in the case of finite-state cdt-MP. Such techniques are extremely powerful and allow for the full solution of those problems, but unfortunately the discrete structure of the finite state and control spaces is crucial, and most of the nice properties MEC has are lost in the more general case of uncountable state spaces.

An alternative approach to the computation of the unbounded-horizon maximal reachability is in [KSL13, Proposition 3], where it is proposed to recast the original fixpoint equation as a linear constrained optimization over the infinite-dimensional space $\mathrm{b}\mathscr{U}(X)$, and to apply numerical methods for its solution. However, the uniqueness of the solution of this problem has not been addressed yet. Other possible alternatives are the theory of Poisson's equations [HLL99, Chapter 7] and the theory of transient cdt-MP [HLL99, Section 9.6], both of which should be applied over the truncated operator ${}_S\mathsf{T}^*$. Another interesting way to approach this problem is to impose the ψ-irreducibility on the model and to tailor the results in [FS02, Chapter 10] developed for the AC performance criterion. All those extensions, however, are out of the scope of the present contribution.

4. Repeated reachability

4.1. Repeated reachability: characterization

It follows from Theorem 1 that model-checking a cdt-MP against any property expressed as a DRA can be reduced to solving the Rabin-like conditions $\Box\Diamond F' \wedge (\neg\Box\Diamond F'')$ over the composition of the cdt-MP with the underlying transition system of the DRA. This result applies in particular to all ω-regular languages and LTL formulae. Unfortunately, we cannot provide a theory that is as comprehensive as for the reachability case (namely, for DFA or safe LTL specifications), as it has been presented in Section 3, and only focus on some partial results. In particular, we focus only on the case of the Büchi acceptance condition $\Box\Diamond F$, which is also easier to characterize by means of its dual property $\Diamond\Box S$, known as *persistence*. As mentioned in Section 2.7, we show how results developed in the setting of gambling theory apply to the cdt-MP case discussed here. Neither the repeated reachability problem nor its dual admit useful bounded-horizon counterparts, so below we omit the symbol ∞ in $\Diamond$ and $\Box$.

Given a cdt-MP $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$, let $S \in \mathscr{B}(X)$ be the set of goal states. A gambling analogue of $\mathfrak{D}$ is given by $\mathfrak{G} = (X, \Gamma)$, where the gambling house defined by

$$\Gamma := \mathsf{proj}_{X\times\mathscr{P}(X)}(\mathsf{Gr}(\mathsf{T}) \cap (\mathbb{K} \times \mathscr{P}(X)))$$

is an analytical subset of $X \times \mathscr{P}(X)$ [BS78, Section 7.6]. It further follows from the equivalence between the cdt-MP and gambling models [Bla76], that we can now invoke results in [MPS91, MS96b] to characterize the value of the repeated reachability problem. In accordance with the mentioned work we call a function

$f \in \mathrm{b}\mathscr{U}(X)$ *excessive*[20] if $\mathsf{T}^* f \leq f$, *deficient* if $(-f)$ is excessive, and *invariant* if its both deficient and excessive. Clearly, invariant functions are precisely the fixpoints of the operator T^*.

The next result provides a characterization of the maximal persistence probability $\mathsf{M}^*(\cdot; \Diamond\Box S)$ and emphasizes its connection with the maximal safety probability $\mathsf{M}^*(\cdot; \Box S)$.

Theorem 7. *For any set $S \in \mathscr{B}(X)$ it holds that $\mathsf{M}^*(\cdot; \Diamond\Box S) \in \mathrm{b}\mathscr{A}^*(X)$. It is also an invariant function, and for any excessive function $f \in \mathrm{b}\mathscr{A}(X)$ satisfying the inequality $f(\cdot) \geq \mathsf{T}^*[\mathsf{M}^*(\cdot; \Box S)]$ it holds that $f(\cdot) \geq \mathsf{M}^*(\cdot; \Diamond\Box S)$. Moreover, the following DP-like recursions hold true:*

$$\mathsf{M}^*(x; \Diamond\Box S) = \lim_{n\to\infty} (\mathsf{T}^*)^n \mathsf{M}^*(x; \Box S), \tag{4.1}$$

where the limit is non-increasing point-wise, for all $x \in X$.

Proof. The result follows immediately from the equivalence of the `cdt`-MP and the gambling models [TA13], where in the latter setting the statement of the theorem is implied by [MPS91, Theorem 1.2] and [MS96b, Theorem 4.5, Corollary 5.5]. □

Note that Theorem 7 connects the maximal safety probability $\mathsf{M}^*(\cdot; \Box S)$ and the maximal persistence probability $\mathsf{M}^*(\cdot; \Diamond\Box S)$. As a result, we can use results on the former function obtained in Section 3 to derive properties of the latter one.

Proposition 7. *For any $S \in \mathscr{B}(X)$: $\mathsf{M}^*(\cdot; \Box S) = 0$ if and only if $\mathsf{M}^*(\cdot; \Diamond\Box S) = 0$.*

Proof. Note that (4.1) immediately implies that $\mathsf{M}^*(\cdot; \Box S) = 0$ is sufficient to claim that $\mathsf{M}^*(\cdot; \Diamond\Box S) = 0$. On the other hand, since $\mathsf{M}^*(\cdot; \Diamond\Box S) \geq \mathsf{M}^*(\cdot; \Box S)$, thanks to Theorem 7 we obtain the converse implication. □

4.2. Repeated reachability: computation

Although the recursions in (4.1) already suggest a possible computational procedure for characterising the value of the maximal probability of persistence $\mathsf{M}^*(\cdot; \Diamond\Box S)$, the scheme requires an infinite number of iterations that are initialized at the maximal safety probability $\mathsf{M}^*(\cdot; \Box S)$, which in turn has to be computed in advance. For the latter quantity we have already discussed non-trivial issues in Section 3: as such, the result of Theorem 7, resorting to finite-horizon approximate computations, is not in general practically applicable. As an alternative, we propose tailoring the technique developed in Theorem 6 to the problem at hand.

Theorem 8. *Let Assumption 1 hold true and further assume stationary policies are sufficient, that is for all $x \in X$ assume that*

$$\mathsf{M}^*(x; \Diamond\Box S) = \sup_{\pi\in\Pi_S} \mathsf{M}^\pi(x; \Diamond\Box S).$$

[20] Note that excessive functions are similar to locally uniformly excessive ones, as in Definition 8.

Suppose that g is a locally π'_0-excessive function on S for some stationary policy $\pi' \in \Pi_S$. Let $E \in \mathscr{B}(X)$ be any open set such that $\inf_{x\in E} \mathsf{M}^*(x;\Diamond\Box S) = 0$, $E \cap \{g \leq 1\} = \emptyset$, E^c is a compact set, and $(E^c)_\infty = S_\infty$. Then for all $x \in X$ and $\varepsilon \in (0,1]$ it holds that

$$|\mathsf{M}^*(x;\Diamond\Box S) - \mathsf{M}^*(x;A_\varepsilon \cup B_\varepsilon)| \leq \max\left(\varepsilon, \sup_{x\in E} \mathsf{M}^*(x;\Diamond\Box S)\right), \tag{4.2}$$

where $B_\varepsilon := \{g \leq \varepsilon\}$ and $A_\varepsilon = (B_\varepsilon \cup E)^c$.

Proof. For any fixed stationary policy we are in the setting of [TA12, Theorem 5], so we are only left with applying Lemma 10. □

Note that provided the requirements of Theorem 8 are met, it is possible to evaluate $\mathsf{M}^*(x;\Diamond\Box S)$ with precise error bounds. Indeed, in such case the set A_ε is compact and simple, hence by Theorem 5 we obtain that $m(A_\varepsilon) < \infty$ and thus the maximal constrained reachability probability $\mathsf{M}^*(x;A_\varepsilon \cup B_\varepsilon)$ can be approximated by the bounded-horizon probabilities. Moreover, the set E here has to be understood as a set where the probability of interest $\mathsf{M}^*(x;\Diamond\Box S)$ is very small, so if one is able to tune E, then the right-hand side in (4.2) can match any given precision. Clearly, the assumptions in Theorem 8 are rather restrictive, and apply only to systems for which the set S serves as a sort of dynamical attractor.

4.3. Comments on the repeated reachability problem

We have mentioned that the characterization in Theorem 7 is taken from the literature on gambling: indeed we have not been able to find similar results obtained for the `cdt`-MP framework. It is interesting to see that the function $\mathsf{M}^*(x;\Diamond\Box S)$ satisfies a fixpoint equation, similarly to the uncontrolled case [TA12]. The connection between the solution of this problem and the value of the maximal safety probability $\mathsf{M}^*(x;\Box S)$ appears to be useful in characterizing simple instances, as we have encountered in Proposition 7.

There is range of literature in gambling on utilities with the form $J := \limsup_{n\to\infty} c(\mathbf{x}_n)$ and $J := \liminf_{n\to\infty} c(\mathbf{x}_n)$, which turn out to be repeated reachability specifications in the case the cost is an indicator function, namely $c(x) = 1_S(x)$. For the lim sup criterion, conditions on sufficiency of stationary policies have been obtained in [Sud69] and [Hil79], while for the lim inf case in [Sud83]. A number of results valid for these criteria are summarized in [MS96a, Section 4], in particular [MS96a, Theorem 9.1, Chapter 4] provides a procedure to find $\mathsf{M}^*(x;\Box S)$ using the transfinite induction algorithm over all countable ordinals, rather than a simple recursion like in (4.1). Although this book only focuses on the case when the state space is countable, some of those results seem to allow for extensions to general Borel state spaces – more research is needed towards this goal. Unfortunately however, they do not seem to lead to practical computational procedures. To the best of our knowledge the result of Theorem 8 is novel, and is an extension of a version for uncontrolled processes in [TA12], where the focus was on studying the stability properties of the absorbing sets. Alternatively, it may be worth invoking some results obtained for recurrence [MT93]: however, such results are only strong when obtained under assumption of ψ-irreducibility of the transition kernel T [FS02,

Chapter 10], which are often restrictive and lead to results that are rarely computational. The AC criterion also seems to be related to the lim sup and lim inf criteria in general, and to the repeated reachability property in particular, however much more research is needed to formally clarify the precise relationship.

To summarize, on the one hand there are many results in gambling related to the repeated reachability problem, however they do not seem to lead to practically useful computational methods. On the other hand, in the `cdt`-MP setting such criteria have not received much attention, and although some related methods for other criteria [FS02] may be useful, such relationship is by no means direct or clear. The current contribution only makes an initial step towards numerical procedures for repeated reachability properties over `cdt`-MP, and much more research on the topic is needed.

5. Case study

In this section the theory developed above is applied to the example presented in Section 2.2, dealing with the control of a power network model. The parameters are chosen as follows: the upper bound for the energy is $M = 2$ and the reserve rate is $c = 0.93$. The consumption of the power plant is assumed to be deterministic with a fixed $p = 0.7$, and the minimal load is fixed to be $\nu_{\min} = 0.8$. The renewable generators are assumed to produce power, the value of which follows a truncated Gaussian distribution with parameters $\mu = 0.1, \sigma = 0.03$ for the first subnetwork, and $\mu = 0.05, \sigma = 0.01$ for the second one, both with support over the interval $[0, 2]$. Similarly, the energy demand follows the same type of distribution, with parameters $\mu = 0.2, \sigma = 0.05$ for the first subnetwork, and $\mu = 0.4, \sigma = 0.07$ for the second one. As a result of this choice of parameters, in practice in the first subnetwork there is less power demand and the renewable generation is more substantial. It is thus expected that the share of the nuclear power plant energy will be higher for the second network: below this intuition is compared with the outputs of the numerical computations.

Let us first resort to the qualitative analysis of the two tasks formulated as automata specifications on Figures 2 and 3. We start by noticing that the cdt-MP we are dealing with is continuous as per Definition 1, so that in particular Theorem 5 can be applied.

With focus on the safety problem (first specification), let the safe set be the square $S = [0.2, 1.5]^2 \subset [0, M]^2$. Clearly, this set is simple in the sense of Definition 7, thus by Theorem 5 the maximal safety probability over the infinite time horizon is equal to 0 over this set. As a consequence, let us now consider a finite horizon $n = 100$ to perform the corresponding computations. The value function is depicted on Figure 5: one can see that even over a relatively long horizon of 100 steps, the safety probability remains equal to 1 over most of the safety set S. Even though the iterations for the safety value function eventually converge to 0 for the infinite time horizon problem, such a convergence is clearly slow. Regarding the optimal policy, we have selected the one at step $n/2 = 50$ as a representative, of which one can see on Figure 6 its u^1-component, namely the fraction of the nuclear plant energy used for the first subnetwork. In particular, whenever the energy level in the first

subnetwork is low the one in the second high, namely $u^1 = 1$, which confirms the intuition that in such a situation all the nuclear power has to be used to maintain the first subnetwork. Conversely, $u^1 = 0$ meaning $u^2 = 1$ over the set where the energy level is high in the first subnetwork and low in the second. In addition, the set $\{u^1 = 0\}$ is larger than $\{u^1 = 1\}$ confirming our intuition that the second network is more fragile (it can rely less on renewable production) and thus requires more energy from the nuclear plant. Finally, Figure 7 which presents the v-component of the policy (total nuclear energy production), and provides a justification for the intuitive idea that for low (high) energy levels v is necessarily high (low).

For the reach-avoid task expressed as the DFA on Figure 3 we can provide a similar analysis. Here we choose the safe set to be $S := [0.2, 2]^2$ and the goal sets $G_1 := (1.8, 2] \times [0.2, 1.8]$, $G_2 := [0.2, 1.8] \times (1.8, 2]$, and finally $G := (1.8, 2]^2$. Again, due to simplicity of set S we obtain that the finite-horizon computations converge exponentially fast to the infinite-horizon value by Theorems 1 and 4. In view of this, as in the case of safety we compute the value function for the finite horizon $n = 100$. The value function on Figure 8 has some intuitive properties: it is equal to 1 on the goal set G, it is equal to 0 over the unsafe set, and is positive elsewhere. The optimal choice of the v-component of the policy is always $v = 1$, as the goal is to maximize the energy level in the two subnetworks: due to this reason, we do not present the trivial plots for the component v. The behaviour of the u^1-component is instead more interesting: we present it at the time step $n/2 = 50$ on Figure 9. One can see that over the safe set, when the energy level in the first subnetwork is high and in the second is low (close to the set G_1), the controller increases the energy level in the first subnetwork ($u^1 \approx 1$). At the same time, after reaching the set G_1 the controller pursues the new goal of maximizing the energy level in the second subnetwork and thus keeps $u^2 \in [0, 0.3]$. Symmetrically, a converse situation holds close to and over the set G_2. Note that on Figure 9 the value of -1 for the policy represents the points where the value does not depend on the control action chosen, which is the case over the goal and the unsafe set.

6. Conclusions

This paper has considered an optimal control synthesis problem, where the probability of a given event is either maximized or minimized over a controlled discrete-time Markov process (`cdt`-MP) model. Using methods from formal languages and automata theory we have proposed a characterization of the events of interest using formulae in linear temporal modal logic (LTL) or derived from deterministic automata. We have extended results known for finite-state `cdt`-MP to general state-space models, and have shown that the original optimal control problems can be reduced to either of two fundamental ones: reachability or repeated reachability. For the former problem, we have provided a full characterization of the dynamic programming (DP) algorithm, and developed a theory of approximation for the unbounded-time problem using computable bounded-horizon counterparts. More restrictive results have been attained for the repeated reachability problem: we have provided a partial characterization od this specification, and proposed a computational technique that can be useful for a class of stable models. We have

further discussed some issues related to the repeated reachability problem: providing a complete answer to them is a promising direction for future research.

[AAP+06] A. Abate, S. Amin, M. Prandini, J. Lygeros, and S. Sastry, *Probabilistic reachability and safe sets computation for discrete time stochastic hybrid systems*, Proceedings of the 45th IEEE Conference of Decision and Control, 2006, pp. 258–263.

[ABFG+93] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus, *Discrete-time controlled Markov processes with average cost criterion: a survey*, SIAM J. Control Optim. **31** (1993), no. 2, 282–344.

[AGLB12] E. Aydin Gol, M. Lazar, and C. Belta, *Language-guided controller synthesis for discrete-time linear systems*, Proceedings of the 15th ACM international conference on Hybrid Systems: Computation and Control (New York, NY, USA), HSCC '12, ACM, 2012, pp. 95–104.

[AKM11] A. Abate, J.-P. Katoen, and A. Mereacre, *Quantitative automata model checking of autonomous stochastic hybrid systems*, Proceedings of the 14th international conference on Hybrid Systems: Computation and Control (New York, NY, USA), HSCC '11, ACM, 2011, pp. 83–92.

[APLS08] A. Abate, M. Prandini, J. Lygeros, and S. Sastry, *Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems*, Automatica **44** (2008), no. 11, 2724–2734.

[AS85] B. Alpern and F. B. Schneider, *Defining liveness*, Information Processing Letters **21** (1985), no. 4, 181 – 185.

[Bel54] R. Bellman, *The theory of dynamic programming*, Bull. Amer. Math. Soc. **60** (1954), 503–515.

[Bel57] ———, *A Markovian decision process*, J. Math. Mech. **6** (1957), 679–684.

[BK08] C. Baier and J.-P. Katoen, *Principles of model checking*, The MIT Press, 2008.

[Bla76] D. Blackwell, *The stochastic processes of Borel gambling and dynamic programming*, Ann. Statist. **4** (1976), no. 2, 370–374.

[Bor91] V. S. Borkar, *Topics in controlled Markov chains*, Pitman Research Notes in Mathematics Series, vol. 240, Longman Scientific & Technical, Harlow, 1991.

[BS78] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control: The discrete time case*, vol. 139, Academic Press, 1978.

[CCL11] D. Chatterjee, E. Cinquemani, and J. Lygeros, *Maximizing the probability of attaining a target prior to extinction*, Nonlinear Analysis: Hybrid Systems **5** (2011), no. 2, 367 – 381.

[CS13] J. Culbertson and K. Sturtz, *A categorical foundation for Bayesian probability*, Applied Categorical Structures (2013), 1–16.

[CY98] C. Courcoubetis and M. Yannakakis, *Markov decision processes and regular events*, IEEE Trans. Automat. Control **43** (1998), no. 10, 1399–1418.

[DAT13] J. Ding, A. Abate, and C.J. Tomlin, *Optimal control of partially observable discrete time stochastic hybrid systems for safety specifications*, Proceedings of the 32nd American Control Conference (2013), 6231–6236.

[Del81] C. Dellacherie, *Capacities and analytic sets*, Cabal Seminar 77-79 (A. S. Kechris, D. A. Martin, and Y. N. Moschovakis, eds.), Lecture Notes in Mathematics, vol. 839, Springer Berlin Heidelberg, 1981, pp. 1–31.

[DPR12] F. Dufour and T. Prieto-Rumeau, *Approximation of Markov decision processes with general state space*, J. Math. Anal. Appl. **388** (2012), no. 2, 1254–1267.

[DS65] L. E. Dubins and L. J. Savage, *How to gamble if you must. Inequalities for stochastic processes*, McGraw-Hill Book Co., New York, 1965.

[Fei83] E. A. Feinberg, *Controlled Markov processes with arbitrary numerical criteria*, Theory of Probability & Its Applications **27** (1983), no. 3, 486–503.

[Fol99] G.B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics, John Wiley & Sons Inc., New York, 1999.

[FS02] E. A. Feinberg and A. Shwartz (eds.), *Handbook of Markov decision processes*, International Series in Operations Research & Management Science, 40, Kluwer Academic Publishers, Boston, MA, 2002, Methods and applications.

[Hil79] T. P. Hill, *On the existence of good Markov strategies*, Trans. Amer. Math. Soc. **247** (1979), 157–176.

[HL89] O. Hernández-Lerma, *Adaptive Markov control processes*, Applied Mathematical Sciences, vol. 79, Springer-Verlag, New York, 1989.

[HLL96] O. Hernández-Lerma and J. B. Lasserre, *Discrete-time Markov control processes*, Applications of Mathematics (New York), vol. 30, Springer Verlag, New York, 1996.

[HLL99] ———, *Further topics on discrete-time Markov control processes*, Applications of Mathematics (New York), vol. 42, Springer-Verlag, New York, 1999.

[JCL+09] S. Jha, E. Clarke, C. Langmead, A. Legay, A. Platzer, and P. Zuliani, *A Bayesian approach to model checking biological systems*, Computational Methods in Systems Biology, Springer, 2009, pp. 218–234.

[Kal97] O. Kallenberg, *Foundations of modern probability*, Probability and its Applications, Springer-Verlag, New York, 1997.

[Kre77a] D. M. Kreps, *Decision problems with expected utility criteria. I. Upper and lower convergent utility*, Math. Oper. Res. **2** (1977), no. 1, 45–53.

[Kre77b] ———, *Decision problems with expected utility criteria. II. Stationarity*, Math. Oper. Res. **2** (1977), no. 3, 266–274.

[Kre78] ———, *Decision problems with expected utility criteria. III. Upper and lower transience*, SIAM J. Control Optim. **16** (1978), 420–428.

[KSL13] M. Kamgarpour, S. Summers, and J. Lygeros, *Control design for specifications on stochastic hybrid systems*, Proceedings of the 16th international conference on Hybrid systems: computation and control (New York, NY, USA), HSCC '13, ACM, 2013, pp. 303–312.

[KV99] O. Kupferman and M. Y. Vardi, *Model checking of safety properties*, Computer aided verification (Trento, 1999), Lecture Notes in Comput. Sci., vol. 1633, Springer, Berlin, 1999, pp. 172–183.

[Mey08] S.P. Meyn, *Control techniques for complex networks*, Cambridge University Press, 2008.

[MPS91] A. Maitra, R. Purves, and W. Sudderth, *A Borel measurable version of König's lemma for random paths*, Ann. Probab. **19** (1991), no. 1, 423–451.

[MS96a] A. P. Maitra and W. D. Sudderth, *Discrete gambling and stochastic games*, Applications of Mathematics (New York), vol. 32, Springer-Verlag, New York, 1996.

[MS96b] ———, *The gambler and the stopper*, Statistics, probability and game theory, IMS Lecture Notes Monogr. Ser., vol. 30, Inst. Math. Statist., Hayward, CA, 1996, pp. 191–208.

[MT93] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Communications and Control Engineering Series, Springer-Verlag London Ltd., London, 1993.

[Par67] K. R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967.

[PP04] D. Perrin and J.-E. Pin, *Infinite words: automata, semigroups, logic and games*, vol. 141, Academic Press, 2004.

[Put94] M.L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*, John Wiley & Sons, Inc., 1994.

[RMT13] M. Rungger, M. Mazo, and P. Tabuada, *Specification-guided controller synthesis for linear systems and safe linear-time temporal logic*, Proceedings of the 16th international conference on Hybrid systems: computation and control (New York, NY, USA), HSCC '13, ACM, 2013, pp. 333–342.

[SB79] S. E. Shreve and D. P. Bertsekas, *Universally measurable policies in dynamic programming*, Math. Oper. Res. **4** (1979), no. 1, 15–30.

[SL95] R. Segala and N. Lynch, *Probabilistic simulations for probabilistic processes*, Nordic Journal of Computing **2** (1995), no. 2, 250–273.

[SL10] S. Summers and J. Lygeros, *Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem*, Automatica **46** (2010), no. 12, 1951–1961.

[Sri98] S. M. Srivastava, *A course on Borel sets*, Graduate Texts in Mathematics, vol. 180, Springer-Verlag, New York, 1998.

[Str66] R. E. Strauch, *Negative dynamic programming*, Ann. Math. Statist. **37** (1966), 871–890.

[Sud69] W. D. Sudderth, *On the existence of good stationary strategies*, Trans. Amer. Math. Soc. **135** (1969), 399–414.

[Sud83] ———, *Gambling problems with a limit inferior payoff*, Math. Oper. Res. **8** (1983), no. 2, 287–297.

[TA12] I. Tkachev and A. Abate, *Stability and attractivity of absorbing sets for discrete-time markov processes*, Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, 2012, pp. 7652–7657.

[TA13] ———, *Formula-free finite abstractions for linear temporal verification of stochastic hybrid systems*, Proceedings of the 16th international conference on Hybrid Systems: Computation and Control (New York, NY, USA), HSCC '13, ACM, 2013, pp. 283–292.

[TA14] I. Tkachev and A. Abate, *Characterization and computation of infinite-horizon specifications over Markov processes*, Theoretical Computer Science **515** (2014), 1 – 18.

[Tab09] P. Tabuada, *Verification and control of hybrid systems: A symbolic approach*, Springer Verlag, New York, 2009.

[TMKA13] I. Tkachev, A. Mereacre, J.-P. Katoen, and A. Abate, *Quantitative automata-based controller synthesis for non-autonomous stochastic hybrid systems*, Proceedings of the 16th international conference on Hybrid Systems: Computation and Control (New York, NY, USA), HSCC '13, ACM, 2013, pp. 293–302.

[Var85] M.Y. Vardi, *Automatic verification of probabilistic concurrent finite state programs*, 26th Annual Symposium on Foundations of Computer Science, 1985, pp. 327–338.

[VW94] M. Y. Vardi and P. Wolper, *Reasoning about infinite computations*, Inf. Comput. **115** (1994), no. 1, 1–37.

[Wol81] P. Wolper, *Temporal logic can be more expressive*, Foundations of Computer Science, 1981. SFCS '81. 22nd Annual Symposium on, 1981, pp. 340–348.

Appendix A.

A.1. Background and notations

A sufficient mathematical background for this paper is encompassed by the following sources: measure theory [Fol99, Chapters 1-3], topology[21] [Fol99, Chapter 4], and basic probability theory [Fol99, Chapter 10]. Below we summarize facts about Borel spaces that are extensively used in the manuscript: most of those we need are covered together with proofs in [BS78, Chapter 7], whereas a more detailed exposition is given in [Par67] and [Sri98].

Given an arbitrary set X we denote by 2^X its powerset, that is the collection of all subsets of X. A complement of $A \subseteq X$ is denoted by $A^c = X \setminus A$. A class $\mathscr{F} \subseteq 2^X$ is called an algebra if it contains the empty set and is closed under finite unions and taking the complement. An algebra $\mathscr{F}$ is called a σ-algebra if in addition it is closed under countable unions. For any class of sets $\mathscr{C} \subseteq 2^X$ we denote by $\sigma(\mathscr{C})$ the smallest σ-algebra that contains $\mathscr{C}$; in that case we say that $\sigma(\mathscr{C})$ is generated by $\mathscr{C}$. In particular, if X is given a topology then $\mathscr{B}(X)$ denotes the Borel σ-algebra of X: the one generated by the class of all open subsets of X. Elements of $\mathscr{B}(X)$ are sometimes referred to as Borel sets. Any topological space by default is assumed to be endowed with its Borel σ-algebra. A topological space X is said to be a (standard) Borel space if it is homeomorphic to a Borel subset of a complete separable metric space. As an example, the set of real numbers $\mathbb{R}$ here is always assumed to be endowed with a usual Euclidian topology, so that $\mathbb{R}$ is a Borel space; all subsets of $\mathbb{R}$ are assumed to be given their inherited subset topologies. As another example, any countable (finite or infinite) set is assumed to be endowed with the discrete topology, which makes it a Borel space.

The set of natural numbers is denoted by $\mathbb{N}$, and we further write $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $\bar{\mathbb{N}}_0 := \mathbb{N}_0 \cup \{\infty\}$. When dealing with ∞ we adopt the following convention: $\infty + 1 = \infty$. For any set X we may use the notation X^ω instead of $X^{\mathbb{N}_0}$. If Y is any other set, we further use the following shorthand notation: $X^\omega \parallel Y^\omega := (X \times Y)^\omega$. Moreover, if $D \subseteq Y^\omega$ we further denote $X \parallel D := \mathsf{proj}_{Y^\omega}^{-1}(D)$ where $\mathsf{proj}_{Y^\omega} : X^\omega \parallel Y^\omega \to Y^\omega$ is an obvious projection map. The latter notation also extends to maps: if Z is some other set and $f : Z \to X^\omega$, $g : Z \to Y^\omega$ are some maps, then $h := f \parallel g : Z \to X^\omega \parallel Y^\omega$ is the unique map such that $\mathsf{proj}_{X^\omega} \circ h = f$ and such that $\mathsf{proj}_{Y^\omega} \circ h = g$. For any set X the identity map on X is given by $\mathrm{id}_X(x) = x$ for all $x \in X$. For $f : X \to Y$ its graph is denoted by

$$\mathsf{Gr}(f) := \{(x, f(x)) : x \in X\} \subseteq X \times Y.$$

All Cartesian products of topological spaces are assumed to be endowed with the corresponding product topologies. In particular, if $(X_k)_{k\in\mathbb{N}}$ is a collection of Borel spaces, and $I \subseteq \mathbb{N}$ then $\mathscr{B}(\prod_{k\in I} X_k) = \bigotimes_{k\in I} \mathscr{B}(X_k)$, i.e. a Borel σ-algebra of a countable product of Borel spaces coincides with the product of their Borel σ-algebras.

[21] For the readers with a background in computer science [PP04, Section 2, Chapter III] may serve as an alternative reference for the introduction to topology.

Given two measurable spaces $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ the map $f : X \to Y$ is said to be measurable if $f^{-1}(\mathscr{Y}) \subseteq \mathscr{X}$; in that case we write $f \in \mathscr{X}/\mathscr{Y}$. If $(Y, \mathscr{Y}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ we simplify the notation and write $f \in \mathscr{X}$ rather than $f \in \mathscr{X}/\mathscr{B}(\mathbb{R})$. If $\mathscr{F}$ is any class of functions $f : X \to \mathbb{R}$ we use $\mathrm{b}\mathscr{F}$ to denote a subclass of bounded functions in $\mathscr{F}$. The class of all bounded functions $\mathrm{b}\mathbb{R}^X$ is assumed to be given a sup-metric $\rho(f, g) := \sup_{x \in X} |f(x) - g(x)|$ which is inherited to all its subclasses. For any two functions $f, g \in \mathbb{R}^X$ we write $\{f \leq g\} := \{x \in X : f(x) \leq g(x)\}$ and similarly for $\{f \geq g\}$ and $\{f = g\}$. We further write $f \leq g$ if and only if $\{f \leq g\} = X$. An important example of functions is given by an indicator function, which for any $A \subseteq X$ is given by

$$1_A(x) := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

For any Borel space X the collection of all probability measures on $(X, \mathscr{B}(X))$ is denoted by $\mathscr{P}(X)$. We always assume the latter to be endowed with a topology of weak convergence, which makes $\mathscr{P}(X)$ a Borel space as well and thus it can be given the Borel σ-algebra $\mathscr{B}(\mathscr{P}(X))$. Given any $A \in \mathscr{B}(X)$ we define an evaluation map $e_A : \mathscr{P}(X) \to [0, 1]$ as $e_A(p) := p(A)$ for any $p \in \mathscr{P}(X)$. It appears that $\mathscr{B}(\mathscr{P}(X))$ is the smallest σ-algebra with respect to which all evaluation maps are measurable. Given a probability measure $p \in \mathscr{P}(X)$ we denote the p-completion of $\mathscr{B}(X)$ by $\mathscr{B}_p(X)$ [BS78]. The universal σ-algebra of a Borel space X is defined as $\mathscr{U}(X) := \bigcap_{p \in \mathscr{P}(X)} \mathscr{B}_p(X)$. For any $p \in \mathscr{P}(X)$ and $f \in \mathrm{b}\mathscr{U}(X)$ we can define the Lebesgue integral $\int_X f \, \mathrm{d}p$ which we also write simply as $p[f]$. If Y is another Borel space, and $g \in \mathscr{U}(X)/\mathscr{B}(Y)$-measurable (i.e. pre-image of every Y-Borel set is universally measurable in X), then any probability measure $p \in \mathscr{P}(X)$ is pushed by g to $g_* p \in \mathscr{P}(Y)$ where the pushforward of the measure is defined by $(g_* p)(A) = p(g^{-1}(A))$ for any $A \in \mathscr{B}(Y)$.

For any two sets X and Y the natural projection from their product onto X is denoted by $\mathsf{proj}_X : X \times Y \to X$ viz. $\mathsf{proj}_X(x, y) = x$ for any $x \in X$ and $y \in Y$. Furthermore, for any $D \subseteq X \times Y$ the x-section of D is defined by

$$D_x := \{y \in Y : (x, y \in D)\}$$

for any $x \in X$. If X is a Borel space, a set $A \subseteq X$ is said to be analytic if there exists $B \in \mathscr{B}(X \times \mathbb{R})$ such that $A = \mathsf{proj}_X(B)$. The collection of all analytic subsets of X is denoted by $\mathscr{S}(X)$. Although it contains the empty set and is closed under countable unions and intersections, it is not closed under taking the complement, so it is not a σ-algebra. The analytical σ-algebra of X is denoted by $\mathscr{A}(X) := \sigma(\mathscr{S}(X))$. It further follows for any Borel space X that

$$\mathscr{B}(X) \subseteq \mathscr{S}(X) \subseteq \mathscr{A}(X) \subseteq \mathscr{U}(X).$$

Given two Borel spaces X and Y we say that a map $f : X \to Y$ is Borel (analytically, universally) measurable if $f \in \mathscr{B}(X)/\mathscr{B}(Y)$ (if $f \in \mathscr{A}(X)/\mathscr{B}(Y)$, if $f \in \mathscr{U}(X)/\mathscr{B}(Y)$). By a stochastic kernel we mean any map of the form $P : X \to \mathscr{P}(Y)$. For any such kernel we write $P(A|x)$ for any $x \in X$ and $A \in \mathscr{B}(Y)$ instead of a more cumbersome version $P(x)(A)$. Moreover, we write $P \in \mathscr{U}(Y|X)$ instead

of $P \in \mathscr{U}(X)/\mathscr{B}(\mathscr{P}(Y))$ and similarly for $\mathscr{A}(Y|X)$ and $\mathscr{B}(Y|X)$. It follows that $P \in \mathscr{B}(Y|X)$ ($\mathscr{A}(Y|X)$, $\mathscr{U}(Y|X)$) if and only if $P(A|\cdot)$ is a Borel (analytically, universally) measurable function for any $A \in \mathscr{B}(Y)$ [Kal97, Lemma 1.37]. The Dirac probability measure at $x \in X$ is denoted by δ_x. Furthermore, for any map $f : X \to Y$ we assign the correspondent kernel δ_f such that $\delta_f(x) := \delta_{\{f(x)\}}$. It follows from [CS13] that f is Borel (analytically, universally) measurable as a map if and only if δ_f is as a kernel.

A function $f : X \to \mathbb{R}$ is said to be lower semi-analytic if $\{f < c\} \in \mathscr{S}(X)$ for any $c \in \mathbb{R}$, and upper semi-analytic if $-f$ is lower semi-analytic. The collection of all lower (upper) semi-analytic functions is denoted by $\mathscr{A}_*(X)$ ($\mathscr{A}^*(X)$). A function $f : X \to \mathbb{R}$ is said to be lower semi-continuous if $\{f \leq c\}$ is closed in X for any $c \in \mathbb{R}$, and upper semi-continuous if $-f$ is lower semi-continuous. The collection of all lower (upper) semi-continuous functions is denoted by $\mathscr{C}_*(X)$ ($\mathscr{C}^*(X)$). The following hierarchy holds for the function classes:

$$\mathscr{C}^*(X),\ \mathscr{C}_*(X) \subseteq \mathscr{B}(X) \subseteq \mathscr{A}^*(X),\ \mathscr{A}_*(X) \subseteq \mathscr{A}(X) \subseteq \mathscr{U}(X).$$

A kernel $P \in \mathscr{B}(Y|X)$ is called continuous if $P : X \to \mathscr{P}(Y)$ is a continuous map. Alternatively, the continuity of the kernel can be characterized as follows: a kernel $P \in \mathscr{B}(Y|X)$ is continuous if and only if $\int_Y f \,\mathrm{d}P \in \mathrm{b}\mathscr{C}^*(X)$ for any $f \in \mathrm{b}\mathscr{C}^*(Y)$.

If (X, ρ_X) and (Y, ρ_Y) are metric space, a map $f : X \to Y$ is called a contraction if there exists a constant $\beta \in [0, 1)$ such that $\rho_Y(f(x'), f(x'')) \leq \beta \cdot \rho_X(x', x'')$ for all points $x', x'' \in X$. The constant β is also called a modulus of a contraction f.

A.2. Auxiliary results

Lemma 9. *Let Y, Y' be arbitrary sets and let $g : Y \to \mathbb{R}$ and $g' : Y' \to \mathbb{R}$ be some functions. Suppose that there exist maps $a : Y \to Y'$ and $a' : Y' \to Y$ such that*

$$g(y) = g'(a(y)), \quad g'(y') = g(a'(y')), \quad \forall y \in Y, y' \in Y'$$

Then: $\inf_{y\in Y} g(y) = \inf_{y'\in Y'} g'(y')$ *and* $\sup_{y\in Y} g(y) = \sup_{y'\in Y'} g'(y')$.

Proof. The following sequences of inequalities

$$\inf_{y\in Y} g(y) = \inf_{y\in Y} g'(a(y)) \geq \inf_{y'\in Y'} g'(y') = \inf_{y'\in Y'} g(a'(y')) \geq \inf_{y\in Y} g(y)$$
$$\sup_{y\in Y} g(y) = \sup_{y\in Y} g'(a(y)) \leq \sup_{y'\in Y'} g'(y') = \sup_{y'\in Y'} g(a'(y')) \leq \sup_{y\in Y} g(y)$$

yield the desired result. □

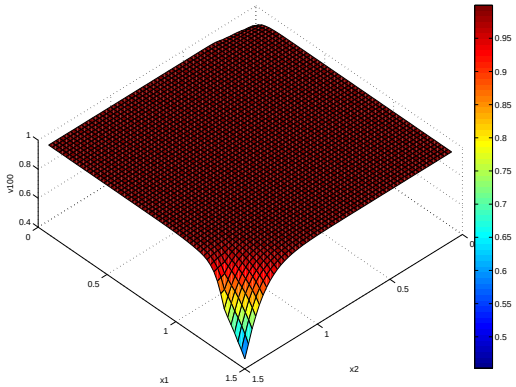
The next lemma shows that point-wise bounds also hold for the optimal values.

Lemma 10. *Let Y be an arbitrary set and consider any two function $f, g : Y \to \mathbb{R}$. If $|f(y) - g(y)| \leq \varepsilon$ for all $y \in Y$ then $|\sup_{y\in Y} f(y) - \sup_{y\in Y} g(y)| \leq \varepsilon$.*

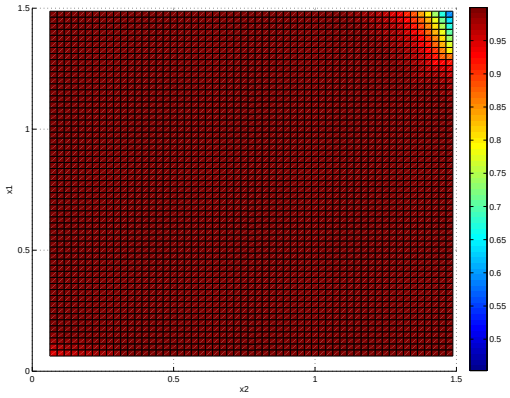
Proof. The proof is given in [HL89, Appendix A.3]. □

Lemma 11. *If Y is a Borel space, the set S is closed in Y and the function $f \in \mathrm{b}\mathscr{C}^*(X)$ is such that $f \geq 0$, then it holds that $1_S \cdot f \in \mathrm{b}\mathscr{C}^*(X)$.*

Proof. Notice that for any $c \leq 0$ it holds that $\{1_S \cdot f \geq c\} = X$, whereas for $c > 0$ we obtain $\{1_S \cdot f \geq c\} = S \cap \{f \geq c\}$ which is a closed set as well. □

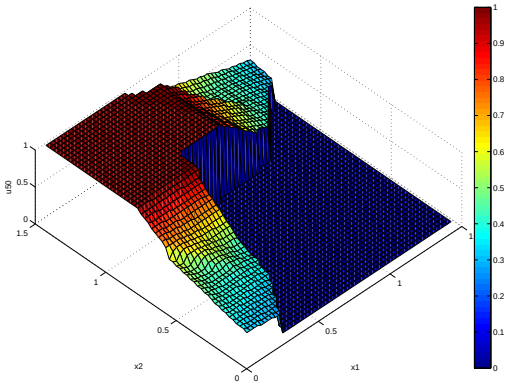


(a) Safety value function plotted in 3d

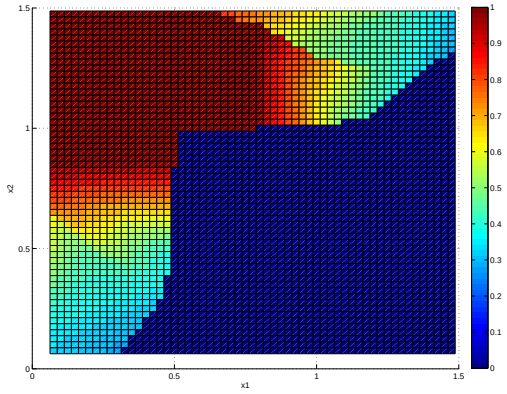


(b) Safety value function plotted in 2d

Figure 5: Safety value function over a finite time horizon of 100 steps.

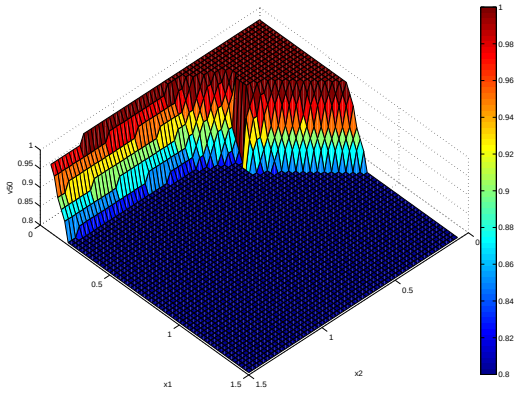


(a) Optimal safety policy, component u plotted in 3d

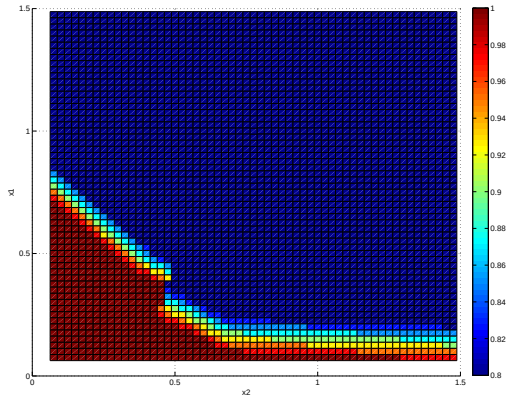


(b) Optimal safety policy, component u plotted in 2d

Figure 6: Optimal safety policy, component u at time step 50.

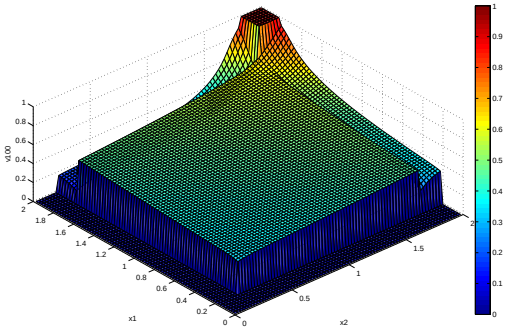


(a) Optimal safety policy, component v plotted in 3d

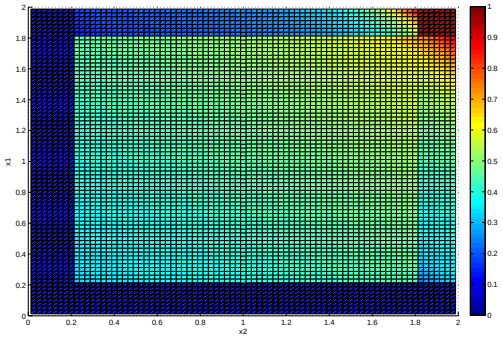


(b) Optimal safety policy, component v plotted in 2d

Figure 7: Optimal safety policy, component v at time step 50.

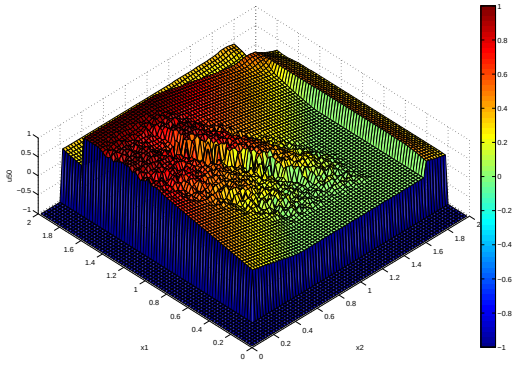


(a) DFA value function plotted in 3d

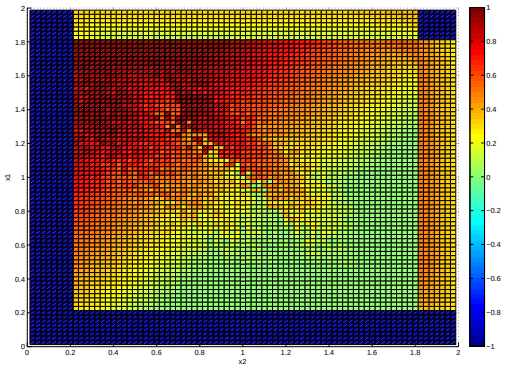


(b) DFA value function plotted in 2d

Figure 8: DFA value function over a finite time horizon of 100 steps.



(a) Optimal DFA policy, component u plotted in 3d



(b) Optimal DFA policy, component u plotted in 2d

Figure 9: Optimal DFA policy, component u at time step 50.