A control Lyapunov function approach for the computation of the infinite-horizon stochastic reach-avoid problem

Ilya Tkachev and Alessandro Abate

***Abstract*— This work is devoted to the solution of the stochastic reach-avoid problem over controlled discrete-time Markov processes (cdt-MP) with general state and action spaces. Whereas the finite time horizon case allows for the use of discretization techniques that compute the quantities of interest with any given precision under mild conditions on the model, the infinite-horizon counterpart demands a more elaborate analysis. This contribution introduces control Lyapunov functions over cdt-MP and shows how these functions help solving the reach-avoid problem over the infinite time horizon. As an example, we show how to apply these technique to the ruin problem arising in the risk theory of insurance companies.**

I. Introduction

General-space controlled discrete-time Markov processes (cdt-MP) provide a rich modeling framework in such application areas as engineering, robotics, system biology, and finance [12], [8], In particular, discrete-time Stochastic Hybrid Systems [**?**] is a subclass of cdt-MP. Such models embed both the probabilistic uncertainty in the system and the presence of a control structure. Most of the classical literature in the cdt-MP theory [6], [9] has been focused on the case of the additive costs, i.e. when at each discrete time step a there is added a cost/reward which depends on the current state of the system and on the implemented control action. Such optimal control problems have been solved using dynamic programming (DP) principles [5]. However, other kinds of performance criteria have not been studied as much in depth.

From a different perspective, research in computer science, and particularly in the field of formal verification, has tackled optimization problems over finite cdt-MP, also known as Markov Decision Processes (MDP). This research has targeted the maximization or the minimization of the likelihood of certain events within the space of trajectories of the process [4, Section 10.6]. Although these optimization objectives do not allow for a direct formulation via additive costs, solution methods have been found for most of these problems. Moreover, by exploiting the finite structure of the model, dedicated software [10], [**?**] has been developed for the numerical solution of optimization problems over MDPs. Unfortunately, such methods are crucially hinging on the discrete structure of MDP, thus further development is required for the case of general cdt-MP.

The stochastic reach-avoid problem targets the quantification of the probability that trajectories of a system reach some goal set G within the state space of the system, while staying within a safe set S before reaching G [**?**]. It represents a generalization of the known probabilistic reachability problem [**?**]: as an example, in the risk theory of insurance companies, ruin probabilities and two-barrier ruin probabilities can be cast as reach-avoid problems [2]. Research on this topic has merged control theory approaches [6] with methods developed within the computer sciences [4], where the reach-avoid problem is known as "constrained reachability." In particular, [**?**] has developed DP for the reach-avoid problem over a finite time horizon and the case of Markov policies. The contribution in [**?**] has in turn considered the infinite-horizon case. The recent work in [16] has further extended the mentioned results and studied quantitative verification of more general properties, such as those expressed via automata. In particular, this work has developed computational methods for the finite time horizon case over history-dependent policies. The infinite time horizon case, however, has been left untouched in [16]: the present work focuses on this framework. We leverage methods developed for the infinite time horizon reach-avoid problem over uncontrolled models [**?**]. In particular, we resort to Lyapunov-like techniques, similar to those used in stochastic stability [11], in order to compute related probabilities.

The structure of the rest of the paper is as follows: Section II introduces the model framework and the reach-avoid problem. To elucidate these concepts, it presents an example taken from risk theory. Section III recapitulates known results on DP for the reach-avoid case, and develops new techniques to tackle the infinite time horizon case. Further, it applies obtained results to the risk theory example. Section IV concludes the work and discusses possible extensions.

II. Preliminaries

A. Notation

Our notation is mostly standard and is inspired by the one in [6, Chapters 7-9]. The set of real numbers is denoted by $\mathbb{R}$, whereas that of natural numbers by $\mathbb{N}$. Furthermore, we denote $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $\bar{\mathbb{N}}_0 := \mathbb{N}_0 \cup \{\infty\}$. Let us emphasize that $\mathrm{b}\mathscr{A}^\circ(X)$ is a collection of bounded lower semi-analytic functions on X, whereas $f \in \mathrm{b}\mathscr{A}^*(X)$ means

This work is supported by the European Commission MoVeS project FP7-ICT-2009-5 257005, by the European Commission Marie Curie grant MANTRAS 249295, by the European Commission NoE HYCON2 FP7-ICT-2009-5 257462, and by the NWO VENI grant 016.103.020.

Ilya Tkachev is with the Delft Center for Systems & Control, Delft University of Technology, The Netherlands. Email: i.tkachev@tudelft.nl. Alessandro Abate is with the Department of Computer Science, University of Oxford, United Kingdom and Delft Center for Systems & Control, Delft University of Technology, The Netherlands. Email: alessandro.abate@cs.ox.ac.uk

$-f \in \mathrm{b}\mathscr{A}^\circ(X)$. We further denote

$$\{f \le c\} = \{x \in X : f(x) \le c\}$$

for any set X, map $f : X \to \mathbb{R}$ and $c \in \mathbb{R}$. A map f between two metric spaces (X, ρ_X) and (Y, ρ_Y) is called a contraction with a modulus $\beta \in [0, 1)$ if for all $x', x'' \in X$

$$\rho_Y(f(x'), f(x'')) \le \beta \cdot \rho_X(x', x'').$$

We follow the convention that $\infty + 1 = \infty$.

B. Models

In this paper we work with `cdt`-MP [16], alternatively known as Markov Control Models [9] or general Markov Decision Processes (MDP) [**?**]. Here we define `cdt`-MP according to the general Borel model studied in [6, Chapters 8,9]. More precisely:

Definition 1. *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* $\mathsf{T} : X \times U \to \mathscr{P}(X)$ *is a Borel measurable stochastic kernel on* X *given* $X \times U$.

Given a `cdt`-MP $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ we say that X is its state space and that U is its action space. Furthermore, $\mathbb{K}_x := \{u \in U : (x, u) \in \mathbb{K}\}$ is the set of actions that are feasible in state $x \in X$, and T is the transition kernel of $\mathfrak{D}$. The behavior of the `cdt`-MP $\mathfrak{D}$ is defined via the following semantics. Consider a time step $k \in \mathbb{N}_0$, and suppose that the current state of $\mathfrak{D}$ is x_k. Depending on the chosen action u_k, the distribution of the next state is given by

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

For example, every stochastic difference equation of the form

$$x_{k+1} = F(x_k, u_k, \xi_k), \tag{2}$$

can be represented via (1) whenever $(\xi_k)_{k\in\mathbb{N}_0}$ is a sequence of iid random variables and F is a Borel measurable map. Moreover, T can be found from F explicitly by

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

for any $B \in \mathscr{B}(X)$, where μ is the distribution of ξ_0. Although any `cdt`-MP allows for a representation as a stochastic difference equation (2), the proof of this fact is not constructive and does not allow for the expression of F given T [9, Section 2.3]. Due to this reason, and for further notational convenience, we choose to deal with the more "general" representation in (1). In the special case of autonomous models, the action space is a singleton $U = \{u\}$ and we call $\mathfrak{D}$ a discrete-time Markov process (`dt`-MP).

To formalize the discussion on behavior of `cdt`-MP, let us introduce the notion of finite and infinite paths:

Definition 2. *Given a* `cdt`*-MP* $\mathfrak{D}$, *an* infinite path *is an infinite sequence*

$$h = (x_0, u_0, x_1, u_1, \dots), \tag{3}$$

where $x_k \in X$ *are the state coordinates and* $u_k \in U$ *are the action coordinates of the path, and* $k \in \mathbb{N}_0$. *The Borel space of all infinite paths* $H = (X \times U)^{\mathbb{N}_0}$, *is called a* canonical sample space *for 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:*

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

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

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 for all $n \in \mathbb{N}_0$:

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

Finite paths h_n are also called *histories* [9], as they represent the past information about the `cdt`-MP, which is further used to synthesize a new action. The above definition of paths allows introducing the notion of control policy.

Definition 3. *Given a* `cdt`*-MP* $\mathfrak{D}$, *a* policy *is a sequence* $\pi = (\pi_n)_{n\in\mathbb{N}_0}$, *where* $\pi_n : H_n \to U$ *is a universally measurable stochastic kernel satisfying*

$$\pi_n(A(x_n)|h_n) = 1. \tag{4}$$

The space of all policies is denoted by Π.

Notice that for some elements of H and H_n it may happen that $u_n \notin U(x_n)$, which reflects the fact that action coordinates may not be feasible: this is allowed for technical reasons and later shown that the corresponding paths are of measure zero. On the other hand, it clearly follows from (4) that any policy $\pi \in \Pi$ can only choose those actions that are are feasible in the current state of the `cdt`-MP. There are several subsets of Π that correspond to the important classes of policies. In case when for all $n \in \mathbb{N}_0$ it holds that $\pi_n(h_n)$ depends only on the current state of the system x_n, we say that $\pi = (\pi_n)_{n\in N_0}$ belongs to the class of Markov policies. The collection of all Markov policies is further denoted by Π_M. Notice that Markov policies do not depend on the whole history, but can be time-dependent. Any Markov policy satisfying $\pi_{n+1}(x) = \pi_0(x)$ for all $n \in \mathbb{N}_0$ and $x \in X$ is called stationary. It shall be clear that stationary policies are exactly Markov, time-independent policies. The collection of all stationary policies is denoted by Π_S.

For any $\pi \in \Pi$ and any initial distribution $\alpha \in \mathscr{P}(X)$, there exists a unique path distribution $\mathsf{P}^\pi_\alpha \in \mathscr{P}(H)$ such that

$$\mathsf{P}^\pi_\alpha(\mathbf{x}_0 \in B) = \alpha(B),$$
$$\mathsf{P}^\pi_\alpha(\mathbf{u}_n \in C|\mathbf{h}_n) = \pi_n(C|\mathbf{h}_n),$$
$$\mathsf{P}^\pi_\alpha(\mathbf{x}_{n+1} \in B|\mathbf{h}_n, \mathbf{u}_n) = \mathsf{T}(B|\mathbf{x}_n, \mathbf{u}_n)$$

for any $n \in \mathbb{N}_0$ and any sets $B \in \mathscr{B}(X)$ and $C \in \mathscr{B}(U)$ [6, Chapter 9]. In the case when $\alpha = \delta_x$ for some state $x \in X$, we simply write P^π_x rather than $\mathsf{P}^\pi_{\delta_x}$.

Finally, let us introduce the following operator that acts on the class $\mathscr{U}(X)$: for any $\pi \in \Pi_S$ we define

$$\mathsf{T}^\pi f(x) := \int\limits_{\mathbb{K}_x \times X} f(y)\mathsf{T}(\mathrm{d}y|x,u)\pi_0(\mathrm{d}u|x),$$

and for the special case $\pi(x) = \delta_{\{u\}}(\cdot)$ we write T^u.

C. Stochastic reach-avoid problem

A reach-avoid problem [?] combines two requirements over a trajectory of a process:

i. the trajectory of the process has to reach some goal set G in a finite time;
ii. the trajectory of the process has to stay within the safe set S until the moment it reaches G.

In case the process is deterministic, the goal of the deterministic reach-avoid problem is to find a control policy that makes the trajectory satisfy i. and ii. However, in case the dynamics of the process is stochastic (as in case of `cdt`-MP), in general there is no control policy that assures the satisfaction of these constraints for any possible realization of the process. On the other hand, one can talk about the probability that the trajectory of a stochastic process satisfies the desired property. Due to this reason, the task of the stochastic reach-avoid problem can be formulated as finding a policy that maximizes (or, conversely, minimizes) such a probability [?], [?]. The special case when S is the whole state space, and thus only constraint i. has to be satisfied, is known as *reachability* of the goal set G [3]: for example, the stochastic reachability problem was addressed in [?]. Here we treat reachability as a special case of reach-avoid.

Clearly, the stochastic reach-avoid problem is quantitative in its nature in the sense that it is characterized by probabilities lying in the interval $[0, 1]$, rather than by binary outcomes `true` or `false`. It is thus convenient to formally introduce the stochastic reach-avoid problem by means of the corresponding value function [13]. Such a function, given the initial state of the process x, the safe and goal sets S and G, and a time horizon n, gives a probability that the trajectory of the process satisfies the reach-avoid property within n steps.

More precisely, let $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ be some arbitrary `cdt`-MP. For any set $A \subseteq X$ let the random variable

$$\tau_A(h) := \inf\{k \geq 0 : \mathbf{x}_k(h) \in A\}$$

be the first hitting time of A. Clearly, $\tau_A : H \to \bar{\mathbb{N}}_0$ is a Borel-measurable map whenever $A \in \mathscr{B}(X)$. Hence, for any two sets $S, G \in \mathscr{B}(X)$ it holds that

$$W_n^\pi(x; S, G) := \mathsf{P}_x^\pi(\tau_G \leq \tau_{S^c}, \tau_G \leq n), \quad n \in \mathbb{N}_0,$$
$$W_\infty^\pi(x; S, G) := \mathsf{P}_x^\pi(\tau_G \leq \tau_{S^c}, \tau_G < \infty),$$

are well-defined universally measurable functions for any policy $\pi \in \Pi$. We call functions $W_{(\cdot)}^\pi$ reach-avoid value functions, and further say that W_n^π is a finite-horizon value function if $n \in \mathbb{N}_0$, whereas W_∞^π is the infinite-horizon one.

Notice that directly from the definition of functions W_n^π it follows that

$$W_n^\pi(x; S, G) = W_n^\pi(x; S \setminus G, G)$$

for any sets $S, G \in \mathscr{B}(X)$, time horizon $n \in \bar{\mathbb{N}}_0$, policy $\pi \in \Pi$ and initial state $x \in X$. Due to this reason, without loss of generality we often assume that the safe set S and the goal set G are disjoint.

Finally, we are able to introduce the optimization problems that are the goal of this paper. For any $S, G \in \mathscr{B}(X)$, $n \in \bar{\mathbb{N}}_0$ and $x \in X$ we define the following quantities:

$$W_n^*(x; S, G) := \sup_{\pi \in \Pi} W_n^\pi(x; S, G), \tag{5}$$
$$W_n^\circ(x; S, G) := \inf_{\pi \in \Pi} W_n^\pi(x; S, G). \tag{6}$$

We further say that the policy $\pi^* \in \Pi$ is optimal for (5) if it holds that $W_n^*(\cdot; S, G) = W_n^{\pi^*}(\cdot; S, G)$. An optimal policy π° for Problem (6) is defined similarly. Note that optimal policies do not have to be unique, which is in particular easy to show for the reach-avoid problem. Clearly, whenever the trajectory of the process reaches $G \cup S^c$, it does not have further affect on the satisfaction of the reach-avoid property. Thus, $W_n^{\pi'}(\cdot; S, G) = W_n^{\pi''}(\cdot; S, G)$ for any $n \in \bar{\mathbb{N}}_0$ whenever policies $\pi', \pi'' \in \Pi$ agree on $S \setminus G$.[1]

We are interested in solving problems (5), (6), namely in finding the optimal value functions and possibly also constructing optimal policies that realize such functions. However, the explicit solution is not possible to find in the general case [?], [?]. Due to this reason, our goal is to provide methods to compute quantities of interest with any given precision.

D. Example: the ruin of an insurance company

To elucidate concepts defined above, let us consider the following example of a stochastic reach-avoid problem important in risk theory – the ruin problem [2]. We introduce it over a particular model of a controlled risk process that was studied in [7].

The ruin problem is focused on the probability that the capital of the insurance company becomes negative at some moment in time. Consider $X := \mathbb{R} \times \mathbb{I}$ to be the state space of the model, where $\mathbb{I}$ is a countable set. Here the $\mathbb{R}$-coordinate represents the capital, and the $\mathbb{I}$-coordinate represents the short-term interest rate. Furthermore, let $u_{\min} \in (0, 1]$ be the minimal retention level, and define $U := [u_{\min}, 1]$ to be the space of actions. We also assume that all actions are feasible regardless of the current state, that is $\mathbb{K}_x = U$ for all $x \in X$. We are only left to describe the transition kernel.

Instead of writing the kernel T explicitly, let us characterize the dynamics of the system by a stochastic difference equation as in (2). Let r_k be the capital of the company at time $k \in \mathbb{N}_0$, and let i_k be the short-term interest rate at that time, then $x_k = (r_k, i_k) \in \mathbb{R} \times \mathbb{I}$. Let i_k evolve independently of r_k as a Markov chain with a stochastic matrix $P = (p_{ij})_{i,j \in \mathbb{I}}$, and let r_k be updated as

$$r_{k+1} = r_k(1 + i_k) + c - (1 + \theta)(1 - u_k)\bar{\xi} - u_k \xi_k,$$

where $(\xi_k)_{k \in \mathbb{N}_0}$ is a sequence of non-negative iid random variables with a distribution μ, which represent the total

[1] A related discussion is also given in [?].

insurance claims during each time period $k \in \mathbb{N}_0$, and $\bar{\xi} = \mathsf{E}\xi_0$ it their expected value. The parameters are: c, the constant premium, and θ, the added safety loading satisfying

$$c - (1+\theta)(1-u_{\min})\bar{\xi} \geq 0. \tag{7}$$

The model can be explained as follows: at each time step an insurance company receives a premium c from the clients and has to pay claims ξ_k back to the clients in insurance cases. The claims are further reinsured, thus the company can decide to pay only a fraction $u_k \in [u_{\min}, 1]$ of the total claim, whereas the remaining part is paid to the client by the reinsurer. Due to this reason, the insurance company further pays a premium $(1+\theta)(1-u_k)\bar{\xi}$ to the reinsurer. As a result, condition (7) means that the premiums received by the insurance company from its clients are greater than premiums paid by the insurance company to the reinsurer. The detailed discussion on the model can be found in [7] and references therein. Finally, the ruin probability can be defined as a reach-avoid problem over sets $S = X$, and $G = (-\infty, 0) \times \mathbb{I}$, which is equivalent to the classical definition given in [2], [7]. Clearly, the main focus in the risk theory in case the control structure is presented is in minimizing (rather than maximizing) the ruin probability. We address this task in Section III-C below.

III. Solution of the stochastic reach-avoid problem

A. Dynamic programming characterization

As we have mentioned above, the stochastic reachability problem over a cdt-MP $\mathfrak{D} = (X, U, \mathbb{K}, \mathsf{T})$ can be considered as a special case of the reach-avoid one, where the safe set is $S = G^c$. Interestingly, it appears that the reach-avoid problem can be considered as a reachability one over a modified cdt-MP, where the avoid set S^c is forced to be invariant under the dynamics of the cdt-MP – see e.g. [16, Section 3.1]. This fact is very important here as it allows for the characterization of the reach-avoid value functions W by means of DP.

A known issue of stochastic reachability and of the reach-avoid problem is that their value functions do not allow in general for the additive cost formulation [6], [9], where the theory is quite rich. Due to this reason, different cost formulation have been proposed in the literature: a multiplicative one [**?**], a sum-multiplicative one [**?**], and an additive up to the hitting time [**?**]. Such formulations allow to obtain a part of the results for the additive cost formulation, but in some cases the required conditions are quite conservative.[2] To cope with these issues, the work in [16] has proposed an equivalent additive cost formulation for the reachability over an augmented cdt-MP, derived DP recursions for the finite time horizon, and DP fixpoint equations for the infinite time horizon. Since the reach-avoid problem was proved to have an equivalent reachability formulation, these results can apply to the present case. Below we shortly summarize them.

From now on we assume sets $S, G \in \mathscr{B}(X)$ to be fixed and disjoint. We introduce the following operators:

$$\mathfrak{T}^* f(x) := \sup_{u \in \mathbb{K}_x} \int_X f(y) \mathsf{T}(\mathrm{d}y|x,u), \quad f \in \mathrm{b}\mathscr{A}^*$$

$$\mathfrak{T}^\circ f(x) := \inf_{u \in \mathbb{K}_x} \int_X f(y) \mathsf{T}(\mathrm{d}y|x,u), \quad f \in \mathrm{b}\mathscr{A}^\circ$$

Given two disjoint sets $S, G \in \mathscr{B}(X)$ we further denote

$$\mathfrak{R}^* f(x) := 1_G(x) + 1_S(x)\mathfrak{T}^* f(x), \quad f \in \mathrm{b}\mathscr{A}^*$$

$$\mathfrak{R}^\circ f(x) := 1_G(x) + 1_S(x)\mathfrak{T}^\circ f(x), \quad f \in \mathrm{b}\mathscr{A}^\circ.$$

It follows from the discussion in [6, Section 8.2] that all four operators map their domains into themselves. The following result characterizes the reach-avoid value functions via DP recursions.

Proposition 1. *It holds that functions W_n^* and W_n° belong to classes $\mathrm{b}\mathscr{A}^*(X)$ and $\mathrm{b}\mathscr{A}^\circ(X)$ respectively, and that*

$$W_{n+1}^* = \mathfrak{R}^*[W_n^*], \qquad W_{n+1}^\circ = \mathfrak{R}^\circ[W_n^\circ]. \tag{8}$$

for any $n \in \bar{\mathbb{N}}_0$, where $W_0^* = W_0^\circ = 1$.

It is now possible to provide methods for the solution of the stochastic reach-avoid problems. There is a strong distinction between the finite-horizon case, which allows for the solution under mild assumptions, and the infinite-horizon one, where more elaborate analysis is required.

B. Infinite time horizon reach-avoid

In the theory of optimal control, and in particular in the area of DP, the class of discounted problems refers to the case when the Bellman DP operator [5] is contractive. Such a property has some nice consequences: the infinite-horizon cost is the unique fixpoint of this operator, and also can be efficiently approximated by means of the finite-horizon costs, as it follows from the contraction mapping theorem [14]. This approach is interesting, since the finite-horizon reach-avoid problem can be precisely computed under certain assumptions on the dynamics of the model. The computational details can be found in [16, Section 4].

In the present context, the reach-avoid problem [16] is never discounted, so that the analysis becomes more subtle. In particular, (8) for $n = \infty$ are fixpoint equations for operators $\mathfrak{R}^*$ and $\mathfrak{R}^\circ$, whose solutions are not unique in general. Indeed, for the case of reachability $S = G^c$ it holds that $f \equiv 1$ always solves both equations, although clearly the solution of the reachability problem in general may be different from the constant one. Moreover, it follows that the uniqueness of fixpoints of r^* or $\mathfrak{R}^\circ$ in the case of reachability implies a trivial constant optimal reachability. As a result, any non-trivial case of the reachability is characterized by the non-unique solutions of the corresponding fixpoint equations.[3]

[2] As an example, for the DP fixpoint equation [**?**] required that the first hitting time of $S^c \cup G$ is finite P^π-a.s. under any Markov policy π, which in the case of reachability corresponds to the constant solution $W_\infty^* \equiv 1$ [16]. Thus, such result is not relevant for the reachability problem being a special case of the reach-avoid one.

[3] The autonomous case even allows for the sharper statements that relate uniqueness, triviality and contractivity – see e.g. [**?**].

Although the corresponding additive cost problem is not discounted, still the operators $\mathfrak{R}^*$ or $\mathfrak{R}^\circ$, related to the reach-avoid problem, can be contractive depending on the set S. However, the discussion above motivates looking into the non-contractive case as well. Due to this reason, we first focus on the former (simpler) case and further show when the latter case can be reduced to the former one.

Theorem 1. *If for some sets $S, G \in \mathscr{B}(X)$ and some integer $m \in \mathbb{N}_0$ it holds that $(\mathfrak{R}^*)^m$ (or $(\mathfrak{R}^\circ)^m$) are contractions on $\mathrm{b}\mathscr{A}^*(X)$ (or $\mathrm{b}\mathscr{A}^\circ(X)$) with a modulus $\beta \in [0,1)$, then $\mathfrak{R}^*$ (or $\mathfrak{R}^\circ$) has a unique fixpoint and*

$$|W^*_\infty(x;S,G) - W^*_{mn}(x;S,G)| \le \beta^n$$
$$(\text{or } |W^\circ_\infty(x;S,G) - W^\circ_{mn}(x;S,G)| \le \beta^n),$$

for any $n \in \mathbb{N}_0$ and any $x \in X$.

Remark 1. *Contractivity of the operators $(\mathfrak{R}^*)^m$ and $(\mathfrak{R}^\circ)^m$ is hard to verify since they are non-linear. It is thus of interest to look for general sufficient conditions assuring contractivity, else it may be possible to verify contractivity over a particular given problem where the shape of the kernel T and of the sets S, G are known.*

*The lack of contractivity, on the other hand, may be possibly related to the concepts of weakly and strongly absorbing sets [16], as much as it is the case for autonomous models [**?**].*

The result of Theorem 1 shows that in case some power of the DP operator for the reach-avoid problem is contractive, it is possible to reduce the infinite-horizon problem to the finite-horizon one. Our next step is to provide a method to reduce the general, non-contractive case to a related contractive one, with guarantees on the error introduced by such a reduction. However, for it we need to assume that stationary policies are sufficient for the optimization of the reach-avoid problem on the infinite time horizon.

Assumption 1. *The following identities hold true:*

$$W^*(x;S,G) = \sup_{\pi\in\Pi_S} W^\pi(x;S,G),$$
$$W^\circ(x;S,G) = \inf_{\pi\in\Pi_S} W^\pi(x;S,G).$$

Some sufficient conditions for Assumption 1 to hold can be found in [6, Chapter 9].[4]

Theorem 2. *Let $S, G \in \mathscr{B}(X)$ be any two disjoint sets. Further, let $C \in \mathscr{B}(X)$ be any subset of S such that for some $m \in \mathbb{N}_0$, $(\mathfrak{R}^*_{S',G})^m$ and $(\mathfrak{R}^\circ_{S',G})^m$ are contractions on $\mathrm{b}\mathscr{A}^\circ(X)$ with a modulus β, where $S' := S \backslash C$. If Assumption 1 holds true then*

$$|W^*_\infty(\cdot;S',G) - W^*_\infty(\cdot;S,G)| \le \chi^*_C, \quad \forall x \in X \tag{9}$$
$$|W^\circ_\infty(\cdot;S',G) - W^*_\infty(\cdot;S,G)| \le \chi^\circ_C, \quad \forall x \in X \tag{10}$$

[4] Although such conditions are formulated over additive cost functions, they clearly can be used in the present framework thanks to the equivalence between reach-avoid and additive cost as in [16].

where constants χ^*_C and χ°_C are given by

$$\chi^*_C := \sup_{\pi\in\Pi_S}\left(\sup_{y\in C} W^\pi_\infty(y;S,G)\right),$$
$$\chi^\circ_C := \inf_{\pi\in\Pi_S}\left(\sup_{y\in C} W^\pi_\infty(y;S,G)\right).$$

Notice that under the assumptions of Theorem 2, the reach-avoid problem for sets S' and G falls into the class of contractive problems and thus can further be solved by means of Theorem 1. This motivates looking into the error bounds given by (9) and (10), and in particular providing non-trivial estimations of χ^*_C and χ°_C. Recall that the latter quantities, ought to be used in computations of value functions W^*_∞ and W°_∞, are defined using these value functions. We obtain such estimations using appropriate Lyapunov-like functions.

Definition 4. *Given a policy $\pi \in \Pi_S$, a Borel-measurable function $g : X \to [0,\infty)$ is called π-locally excessive on the set S if $\{g \le 1\} \subseteq S$, and $g(x) \le 1$ implies $\mathsf{T}^\pi g(x) \le g(x)$.*

A Borel-measurable function $g : X \to [0,\infty)$ is called uniformly locally excessive on the set S if $\{g \le 1\} \subseteq S$, and $g(x) \le 1$ implies $\mathsf{T}^u g(x) \le g(x)$ for all $u \in \mathbb{K}_x$.

Remark 2. *In definition above, one can replace the threshold 1 with the existence of a positive constant δ. Note however, that in such a case the defined objects are invariant under the scaling on the positive constant, so that δ can be initially chosen to be 1, as it is in our case.*

Note that π-locally excessive functions are generalization of excessive functions [15] and locally excessive functions [11], [**?**]. They can be thought of as controlled Lyapunov functions for `cdt`-MP: there exists a policy π, which leads the value of such a function to decrease along the dynamics of the process. Uniformly locally excessive functions, in turn, can be considered as the analogue of Lyapunov functions used in the stability analysis of non-stochastic differential inclusions [1]: their behavior is decreasing along the trajectory of the system regardless of the chosen policy, as the following lemma shows.

Lemma 1. *If g is a uniformly excessive function for S, then $\mathsf{T}^\pi g(x) \le g(x)$ for any $x \in \{g \le 1\}$ and any $\pi \in \Pi_S$.*

We are now ready to relate χ^*_C and χ°_C to the corresponding locally excessive functions.

Theorem 3. *Let g^* be a uniformly locally excessive function for the set S, and let g° be a $\hat\pi$-locally excessive function for set S, and for some policy $\hat\pi \in \Pi_S$. Further, define sets $C^*(\varepsilon) := \{g^* \le \varepsilon\}$ and $C^\circ(\varepsilon) := \{g^\circ \le \varepsilon\}$. Then for any $\varepsilon \le 1$ it holds that $\chi^*_{C^*(\varepsilon)} \le \varepsilon$ and $\chi^\circ_{C^\circ(\varepsilon)} \le \varepsilon$.*

Combining the results of theorems 1, 2 and 3, we obtain that the knowledge of the appropriate locally excessive functions allows reducing the general non-contractive problem to a new contractive one, with an explicit bound on the error. The presented results extend those for the autonomous models summarized in [**?**]. Unfortunately, even if locally excessive functions are known, in general case one additionally has to asses the contractivity of DP operators $\mathfrak{R}^*_{S\setminus C^*(\varepsilon),G}$ and $\mathfrak{R}^\circ_{S\setminus C^\circ(\varepsilon),G}$. Let us further mention that in the case of autonomous models with certain continuity properties of the kernels [?], one can assure the contractivity of these operators due to their linearity. In the general case of cdt-MP, however, each case needs to be studied separately.

C. Application to the ruin problem

To enlighten our approach, let us briefly summarize how does it apply to the case or ruin problem over the model introduced in Section II-D. First of all, notice that recursions and fixpoint equations derived in [7, Lemma 2] are given for a fixed stationary policy, and hence are covered in the literature on autonomous models [13]. Here, instead, Proposition 1 provides recursions and fixpoint equations for the optimal solution of the controlled ruin problem, in particular for finding the minimal ruin probability.

Let us denote $C(\varepsilon) := [\varepsilon^{-1}, \infty)$ for any $\varepsilon \geq 1$, then since $u_{\min} > 0$ and if μ has an unbounded support, one can show that the operator $\left(\mathfrak{R}^\circ_{S\setminus C(\varepsilon),G}\right)^2$ is a contraction. Hence, it is possible to compute the value function $W^\circ_\infty(\cdot; S\setminus C(\varepsilon), G)$ with any given precision using Theorem 1.

Let us further assume that there exists $\hat{u} \in U$ such that

$$\hat{u}\bar{\xi} < c - (1+\theta)(1-\hat{u})\bar{\xi}, \tag{11}$$

so the expected amount of claims the company has to pay is lower than its revenue. The case $\hat{u} = 1$ in (11) corresponds to the Net Profit Condition (NPC) [2], which means that the capital of the company grows on average.

Under condition (7) define $\hat{\pi} \equiv \hat{u}$ to be a constant stationary policy. It follows from [7, Section 2] that

$$\sup_{y\in C(\varepsilon)} W^{\hat{\pi}}_\infty(y; S, G) \leq \exp(-R_0\varepsilon^{-1}),$$

where the constant R_0 is given in [7, Equation (10)]. As a result, $\chi^\circ_{C(\varepsilon)} \leq \exp(-R_0\varepsilon^{-1})$ and thus it can be made arbitrary small for sufficiently small ε. Hence, using Theorem 2 we can reduce the evaluation of the original minimal ruin probability $W^\circ_\infty(\cdot; S, G)$ to that of the $W^\circ_\infty(\cdot; S\setminus C(\varepsilon), G)$, the computability of which has been already discussed above, with an error that is as small as needed. The actual computations, however, go beyond the scope of this contribution.

IV. Conclusions

This contribution has focused on the optimization problem of reach-avoid probabilities over a general state-space, controlled discrete-time Markov process (cdt-MP). A special attention has been paid to the infinite-horizon case, as the finite-horizon one is known to allow for computational techniques [16]. We have shown that in case of contractive operators the reach-avoid problem can be solved with any precision, and further provided a method to reduce the general case to the contractive one under the assumed knowledge of Lyapunov-like locally excessive functions. Finally, we have elaborated on the developed technique over the example of a ruin problem taken from the risk theory.

Note that there are some questions that are still left open. In particular, for the computational direction it is important to work out sufficient conditions of the contractivity for Bellman operators for the reach-avoid problem. As an example, it is interesting to study the connection between contractivity and level sets of locally excessive functions, as done over autonomous models.

References

[1] D. Angeli, B. Ingalls, E. D. Sontag, and Y. Wang. Uniform global asymptotic stability of differential inclusions. *J. Dynam. Control Systems*, 10(3):391–412, 2004.

[2] S. Asmussen. *Ruin probabilities*, volume 2 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific Publishing Co. Inc., River Edge, NJ, 2000.

[3] J.-P. Aubin. *Viability theory*. Springer, 2009.

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

[5] R. Bellman. A Markovian decision process. *J. Math. Mech.*, 6:679–684, 1957.

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

[7] M. A. Diasparra and R. Romera. Bounds for the ruin probability of a discrete-time risk process. *Journal of Applied Probability*, 46(1):99–112, 2009.

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

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

[10] J.-P. Katoen, M. Khattri, and I. S. Zapreev. A Markov reward model checker. In *Proceedings of the Second International Conference on the Quantitative Evaluation of Systems*, QEST '05, pages 243–244, Washington, DC, USA, 2005. IEEE Computer Society.

[11] H. J. Kushner. *Stochastic stability and control*, volume 33. Academic Press, 1967.

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

[13] F. Ramponi, D. Chatterjee, S. Summers, and J. Lygeros. On the connections between PCTL and dynamic programming. In *Proceedings of the 13th ACM international conference on Hybrid Systems: Computation and Control*, pages 253–262, 2010.

[14] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[15] A. Shiryaev, B. Rozovskii, and G. Grimmett. *Optimal stopping rules*. Springer Verlag Berlin Heidelberg, 2008.

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