Parity Objectives in Countable MDPs
(Extended Abstract)

Stefan Kiefer*, Richard Mayr†, Mahsa Shirmohammadi*, Dominik Wojtczak‡
*University of Oxford, UK
†University of Edinburgh, UK
‡University of Liverpool, UK

Markov decision processes (MDPs) are a standard model for dynamic systems that exhibit both stochastic and controlled behavior [18]. The system starts in the initial state and makes a sequence of transitions between states. Depending on the type of the current state, either the controller gets to choose an enabled transition (or a distribution over transitions), or the next transition is chosen randomly according to a defined distribution. By fixing a strategy for the controller, one obtains a probability space of plays of the MDP. The goal of the controller is to optimize the expected value of some objective function on the plays of the MDP. The fundamental questions are "what is the optimal value that the controller can achieve?", "does there exist an optimal strategy, or only ϵ-optimal approximations?", and "which types of strategies are optimal or ϵ-optimal?".

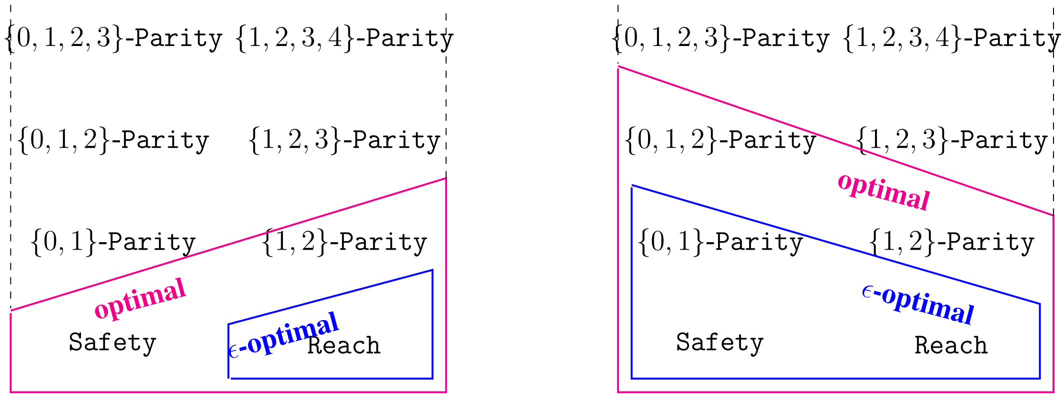
Such questions have been studied extensively for finite MDPs (see e.g. [7] for a survey) and also for certain types of countably infinite MDPs [18], [16]. However, the literature on countable MDPs is mainly focused on objective functions defined w.r.t. numeric costs (or rewards) that are assigned to transitions, e.g. (discounted) expected total reward or limit-average reward. In contrast, we study qualitative objectives that are expressed by Parity conditions and which are motived by formal verification questions.

There are works that studied particular classes of countably infinite, but finitely branching, MDPs that arise from models in automata theory [10], [2], [5], [4], [1]. In each of these papers, a crucial part of the analysis is establishing the existence of optimal strategies of particular structure and memory requirements, but none of them looked at proving such properties for general countable MDPs. Countable MDPs also naturally occur in the analysis of queueing systems [14], gambling [3], and branching processes [17], which have multiple applications. They also show up in the analysis of finite-state models, e.g. in two-player stochastic games [19], [9] when reasoning about an optimal strategy against a fixed (randomised and memory-full) strategy of the opponent.

Finite MDPs vs. Infinite MDPs: It should be noted that many standard properties (and proof techniques) of finite MDPs do *not* carry over to infinite MDPs.

E.g., given some objective, consider the set of all states in an MDP that have nonzero value. If the MDP is finite then this set is finite and thus there exists some minimal nonzero value. This property does *not carry over* to infinite MDPs. Here the set of states is infinite and the infimum over the nonzero values can be zero. As a consequence, even for a reachability objective, it is possible that all states have value > 0, but still the value of some states is < 1. Such phenomena appear already in infinite-state Markov chains like the classic Gambler's ruin problem with unfair coin tosses in the player's favor (0.6 win, 0.4 lose). The value, i.e., the probability of ruin, is always > 0, but still < 1 in every state except the ruin state itself; cf. [11] (Chapt. 14). Another difference is that optimal strategies need not exist, even for qualitative objectives like reachability or parity. Even if some state has value 1, there might not be any single strategy that attains the value 1, but only an infinite family of ϵ-optimal strategies for every $\epsilon > 0$.

Parity objectives: We study general countably infi-



a) Infinitely branching MDPs **b)** Finitely branching MDPs

Fig. 1: For countable MDPs, these diagrams show the memory requirements of optimal and ϵ-optimal strategies for objectives in the Mostowski hierarchy. An objective in a level of the hierarchy subsumes all objectives in lower levels, e.g., $\{0, 1, 2\}$-Parity subsumes $\{1, 2\}$-Parity. We have extended the Mostowski hierarchy to include reachability and safety. The magenta (resp., blue) regions enclose objectives where memoryless deterministic (MD) strategies are sufficient for optimal (resp., ϵ-optimal) strategies; for objectives outside the regions, infinite-memory strategies are necessary. The left diagram is for infinitely branching MDPs; e.g., ϵ-optimal strategies for all but reachability objectives require infinite memory, whereas MD-strategies are sufficient for reachability. The right diagram is for finitely branching MDPs; e.g., optimal strategies (if they exist) can be chosen MD for all objectives subsumed by $\{0, 1, 2\}$-Parity.

nite MDPs with parity objectives. Parity conditions are widely used in temporal logic and formal verification, e.g., they can express ω-regular languages and modal μ-calculus [12]. Every state has a *color*, out of a finite set of colors encoded as natural numbers. An infinite play is winning iff the highest color that is seen infinitely often in the play is even. The controller wants to maximize the probability of winning plays. Subclasses of parity objectives are defined by restricting the set of used colors; these are classified in the Mostowski hierarchy [15] which includes, e.g., Büchi and co-Büchi objectives. Such prefix-independent infinitary objectives cannot generally be encoded by numeric transition rewards as in [18], though both types subsume the simpler reachability and safety objectives.

There are different types of strategies, depending on whether one can take the whole history of the play into account (history-dependent; (H)), or whether one is limited to a finite amount of memory (finite memory; (F)) or whether decisions are based only on the current state (memoryless; (M)). Moreover, the strategy type depends on whether the controller can randomize (R) or is limited to deterministic choices (D). The simplest type MD refers to memoryless deterministic strategies.

The type of strategy needed for an optimal (resp. ϵ-optimal) strategy for some objective is also called the *strategy complexity* of the objective. For finite MDPs, MD-strategies are sufficient for all types of qualitative and quantitative parity objectives [6], [8], but the picture is more complex for countably infinite MDPs.

Since optimal strategies need not exist in general, we consider both the strategy complexity of ϵ-optimal strategies, and the strategy complexity of optimal strategies under the assumption that they exist. E.g., if an optimal strategy exists, can it be chosen MD?

We provide a complete picture of the memory requirements for objectives in the Mostowski hierarchy, which is summarized in Figure 1.

In particular, our results show that there is a strong dichotomy between two different classes of objectives. For objectives of the first class, optimal strategies, where they exist, can be chosen MD. For objectives of the second class, optimal strategies require infinite memory in general, in the sense that all FR-strategies achieve the objective only with probability zero. A similar dichotomy applies to ϵ-optimal strategies. For certain objectives, ϵ-optimal MD-strategies exist, while for all others even ϵ-optimal strategies require infinite memory in general. This is a strong dichotomy because there are no objectives in the Mostowski hierarchy for which other types of strategies (MR, FD, or FR) are both necessary and sufficient. Put differently, for all objectives in the Mostowski hierarchy, if FR-strategies suffice then MD-strategies suffice as well.

We also consider the subclass of countable MDPs that are finitely branching. (Note that these generally still have an infinite number of states.) The above mentioned dichotomies apply here as well, though the classes of objectives where optimal (resp. ϵ-optimal) strategies can be chosen MD are larger than for general countable MDPs.

This extended abstract is based on [13].

REFERENCES

[1] P.A. Abdulla, R. Ciobanu, R. Mayr, A. Sangnier, and J. Sproston. Qualitative analysis of VASS-induced MDPs. In *Proc. of FOSSACS 2016*, volume 9634 of *LNCS*, 2016.

[2] C. Baier, N. Bertrand, and Ph. Schnoebelen. Verifying nondeterministic probabilistic channel systems against omega-regular linear-time properties. *ACM Transactions on Computational Logic*, 9, 2007.

[3] N. Berger, N. Kapur, L. J. Schulman, and V. V. Vazirani. Solvency games. In Ramesh Hariharan, Madhavan Mukund, and V. Vinay, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2008, December 9-11, 2008, Bangalore, India*, pages 61–72, 2008.

[4] T. Brázdil, V. Brožek, K. Etessami, A. Kučera, and D. Wojtczak. One-counter Markov decision processes. In *SODA'10*, pages 863–874. SIAM, 2010.

[5] T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. Reachability in recursive Markov decision processes. *Information and Computation*, 206(5):520–537, 2008.

[6] K. Chatterjee, L. de Alfaro, and T. Henzinger. Trading memory for randomness. In *Proceedings of the First Annual Conference on Quantitative Evaluation of Systems (QEST)*, pages 206–217. IEEE Computer Society Press, 2004.

[7] K. Chatterjee and T. Henzinger. A survey of stochastic ω-regular games. *Journal of Computer and System Sciences*, 78(2):394–413, 2012.

[8] K. Chatterjee, M. Jurdziński, and T. Henzinger. Quantitative stochastic parity games. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '04, pages 121–130, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.

[9] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.

[10] K. Etessami and M. Yannakakis. Recursive Markov decision processes and recursive stochastic games. In *ICALP'05*, volume 3580 of *LNCS*, pages 891–903. Springer, 2005.

[11] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley & Sons, second edition, 1966.

[12] E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics, and Infinite Games*, volume 2500 of *LNCS*, 2002.

[13] S. Kiefer, R. Mayr, M. Shirmohammadi, and D. Wojtczak. Parity Objectives in Countable MDPs. In *Proceedings of LICS 2017 (to appear)*, 2017. Full version available at `https://arxiv.org/pdf/1704.04490.pdf`.

[14] M.Y. Kitaev and V.V. Rykov. *Controlled queueing system*. CRC press, 1995.

[15] A. Mostowski. Regular expressions for infinite trees and a standard form of automata. In *Computation Theory*, volume 208 of *LNCS*, pages 157–168, 1984.

[16] D. Ornstein. On the existence of stationary optimal strategies. *Proc. Am. Math. Soc.*, 20:563–569, 1969.

[17] S.P. Pliska. Optimization of multitype branching processes. *Management Science*, 23(2):117–124, 1976.

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

[19] L.S. Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.