Sure-almost-sure and Sure-limit-sure Window Mean Payoff in Markov Decision Processes

Pranshu Gaba[0009-0000-8012-780X] and Shibashis Guha[0000-0002-9814-6651]

Tata Institute of Fundamental Research, Mumbai, India

Abstract. Given rationals α and β, the *sure-almost-sure* problem for an objective φ in a Markov decision process (MDP) asks if one can simultaneously ensure that all outcomes of the MDP have φ-value at least α (i.e., *sure* α satisfaction), and with probability 1 the outcome has φ-value at least β (i.e., *almost-sure* β satisfaction). The *sure-limit-sure* problem asks if for all $\varepsilon > 0$, one can simultaneously ensure that all outcomes have φ-value at least α, and with probability at least $1 - \varepsilon$ the outcome has φ-value at least β. Moreover, if simultaneous satisfaction of objectives is possible, then one would also like to construct a strategy (for sure-almost-sure) or a family of strategies (for sure-limit-sure) that achieves this.
In this paper, we solve the sure-almost-sure and sure-limit-sure problems for *window mean-payoff objectives*. The window mean-payoff objective strengthens the standard mean-payoff objective by requiring that the average payoff of a finite window that slides over an infinite run be greater than a given threshold. We study two variants of window mean payoff; for both variants we show that sure-almost-sure problem and the sure-limit-sure problem are no harder than sure satisfaction and almost-sure satisfaction when considered separately for these objectives. The full version of this paper is available at [14].

Keywords: Beyond worst-case synthesis · Sure-almost-sure satisfaction · Window mean payoff · Finitary objectives · Markov decision processes

Beyond worst-case synthesis. Classical two-player zero-sum games [1, 15] involve decision-making against a purely antagonistic environment where a threshold performance or a specific behaviour of the system needs to be ensured against *all possible strategies of the environment*, that is even in the worst case. On the other hand, Markov decision processes (MDPs) model uncertainty, and decision-making involves ensuring a *specified behaviour with sufficient probability* or a *high expected performance* against a stochastic environment. Such a stochastic model of the environment, however, does not provide any guarantee on the worst-case performance, and a strategy that is adequate against an adversarial environment may provide a suboptimal performance against a stochastic environment.

In practice, both might be desired simultaneously: A system needs to provide guarantee in the worst-case, as well as ensure some threshold performance with a high probability against a stochastic environment. The beyond worst-case (BWC) framework was introduced in [7] to provide strict worst-case guarantee as well as good expected performance for quantitative specifications.

Window mean payoff. For objectives such as long-run limits of reward functions [13, 16], a play may satisfy the objective while still exhibiting undesired behaviours over arbitrarily long finite segments [9, 10]. Finitary (or window) objectives enforce that undesired behaviour persist only within intervals of bounded length (windows) along the play. In this work, we focus on *window mean-payoff objectives* [9], which are finitary versions of the classical mean-payoff objective. Given a window length $\ell \geq 1$ and a threshold $\lambda \in \mathbb{Q}$, the *fixed window mean-payoff* objective $\mathsf{FWMP}(\ell, \lambda)$ holds if, except for a finite prefix, from every position i in the play there exists a window starting at i of length at most ℓ whose mean payoff is at least λ. The *bounded window mean-payoff* objective $\mathsf{BWMP}(\lambda)$ holds for a play if there exists some $\ell \geq 1$ for which $\mathsf{FWMP}(\ell, \lambda)$ holds for the play.

Related work. In [4], the beyond worst-case setting was studied for qualitative omega-regular objectives encoded as two parity objectives, where the first one needs to be satisfied surely, while the second one needs to be satisfied almost-surely or with a high threshold probability, and the problems were shown to be in $\mathsf{NP} \cap \mathsf{coNP}$; in [11, 12], these problems were studied in the context of stochastic games which are a generalization of MDPs where the environment is both stochastic and adversarial. In [2], Boolean combinations of omega-regular objectives that need to be enforced either surely, almost surely, existentially, or with non-zero probability were studied, and it was shown that both randomization and infinite memory may be required by an optimal strategy. In [3], a combination of parity objective and multiple reachability objectives along with threshold probabilities were considered where the parity objective needs to be satisfied surely and each reachability objective is to be satisfied with the corresponding threshold probability.

Mean-payoff objectives were studied initially in two-player games without stochasticity [13, 16], and finitary versions were introduced as window mean-payoff objectives [9] and have been studied extensively since the last decade. For finitary mean-payoff objectives, the satisfaction problem [6] and the expectation problem [5] were studied in MDPs. Both problems were shown to be in PTIME for the $\mathsf{FWMP}(\ell)$ objective and in $\mathsf{NP} \cap \mathsf{coNP}$ for the BWMP objective.

Contributions. We study two problems for the fixed and bounded window mean-payoff objectives each. Given an MDP $\mathcal{M}$, thresholds $\alpha, \beta \in \mathbb{Q}$:

1. (Sure-almost-sure (SAS)) Does there exist a strategy that ensures (i) a window mean-payoff at least α surely, i.e. against all strategies of an adversarial environment, and (ii) a window mean-payoff at least β almost surely, that is, with probability 1, against a stochastic model of the environment? If yes, then we would also like to synthesize such a strategy.
2. (Sure-limit-sure (SLS)) For every $\varepsilon > 0$, does there exist a strategy that ensures (i) a window mean-payoff at least α surely, and (ii) a window mean-payoff at least β with probability $1 - \varepsilon$? If yes, then we would also like to synthesize such a family of strategies.

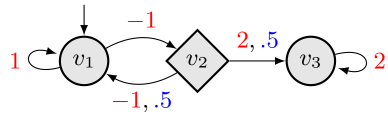

Fig. 1: All vertices are winning for sure-$\mathsf{FWMP}(2,1)$, almost-sure-$\mathsf{FWMP}(2,2)$, and sure-$\mathsf{FWMP}(2,1)$-limit-sure-$\mathsf{FWMP}(2,2)$. However, vertices v_1 and v_2 are not winning for sure-$\mathsf{FWMP}(2,1)$-almost-sure-$\mathsf{FWMP}(2,2)$.

While almost-sure satisfaction and limit-sure satisfaction are equivalent in MDPs [8], we show that sure-almost-sure satisfaction is strictly stronger than sure-limit-sure satisfaction (Figure 1). The same MDP also shows that the winning region for the combination of winning conditions is not merely the intersection of the winning regions for sure-$\mathsf{FWMP}(\ell,\alpha)$ and for almost-sure-$\mathsf{FWMP}(\ell,\beta)$. We have that every vertex is winning for sure-$\mathsf{FWMP}(\ell,\alpha)$ and for almost-sure-$\mathsf{FWMP}(\ell,\beta)$ separately, and yet there exist vertices v_1 and v_2 that are not winning for sure-$\mathsf{FWMP}(\ell,\alpha)$-almost-sure-$\mathsf{FWMP}(\ell,\beta)$. Indeed, we need to make good use of prefix-independence and the finitary nature of window mean-payoff objectives to find the winning region for the combination of winning conditions.

We show that for both $\mathsf{FWMP}(\ell)$ and BWMP, the SAS and SLS problems are no more complex than solving sure satisfaction [9] or almost-sure satisfaction [6] of the same objective separately. We also show that the memory requirement for winning strategies may be higher for sure-almost-sure satisfaction compared to satisfying the objective either surely or almost surely. Our results are summarized in Table 1.

Table 1: Results (ours shaded) for window mean-payoff satisfaction in MDPs

Winning condition		Complexity	Memory bound: lower	Memory bound: upper
$\mathsf{FWMP}(\ell)$	sure [9]	PTIME	$\ell-1$	ℓ
	almost-sure [6]	PTIME	$\ell-1$	ℓ
	SAS	PTIME	$\Omega(\max\{\lvert V\rvert,\ell\})$	$\mathcal{O}(\lvert V\rvert\cdot\ell)$
	SLS	PTIME	$\Omega\left(\frac{1-\varepsilon}{(\mathbb{P}_{\min})^{\lvert V\rvert}}+\ell\right)$	$\mathcal{O}\left(\frac{\lvert V\rvert\cdot\log(1/\varepsilon)}{(\mathbb{P}_{\min})^{\lvert V\rvert}}+\ell\right)$
BWMP	sure [9]	NP ∩ coNP	memoryless	memoryless
	almost-sure [6]	NP ∩ coNP	memoryless	memoryless
	SAS	NP ∩ coNP	$\Omega(\lvert V\rvert\cdot w_{\max})$	$\mathcal{O}(\lvert V\rvert^3\cdot\lvert w_{\max}\rvert)$
	SLS	NP ∩ coNP	$\Omega\left(\frac{1-\varepsilon}{(\mathbb{P}_{\min})^{\lvert V\rvert}}\right)$	$\mathcal{O}\left(\frac{\lvert V\rvert\cdot\log(1/\varepsilon)}{(\mathbb{P}_{\min})^{\lvert V\rvert}}\right)$

Acknowledgments. We acknowledge support of the Department of Atomic Energy, Government of India, under project no. RTI4014.

References

1. Apt, K.R., Grädel, E.: Lectures in Game Theory for Computer Scientists. Cambridge University Press (2011). https://doi.org/10.1017/CBO9780511973468
2. Berthon, R., Guha, S., Raskin, J.: Mixing Probabilistic and non-Probabilistic Objectives in Markov Decision Processes. In: LICS. pp. 195–208. ACM (2020). https://doi.org/10.1145/3373718.3394805
3. Berthon, R., Katoen, J., Winkler, T.: Markov Decision Processes with Sure Parity and Multiple Reachability Objectives. In: RP. Lecture Notes in Computer Science, vol. 15050, pp. 203–220. Springer (2024). https://doi.org/10.1007/978-3-031-72621-7_14
4. Berthon, R., Randour, M., Raskin, J.: Threshold Constraints with Guarantees for Parity Objectives in Markov Decision Processes. In: ICALP. LIPIcs, vol. 80, pp. 121:1–121:15 (2017). https://doi.org/10.4230/LIPIcs.ICALP.2017.121
5. Bordais, B., Guha, S., Raskin, J.: Expected Window Mean-Payoff. In: FSTTCS. LIPIcs, vol. 150, pp. 32:1–32:15 (2019). https://doi.org/10.4230/LIPIcs.FSTTCS.2019.32
6. Brihaye, T., Delgrange, F., Oualhadj, Y., Randour, M.: Life is Random, Time is Not: Markov Decision Processes with Window Objectives. Logical Methods in Computer Science **Volume 16, Issue 4** (Dec 2020). https://doi.org/10.23638/LMCS-16(4:13)2020
7. Bruyère, V., Filiot, E., Randour, M., Raskin, J.F.: Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games. Information and Computation **254**, 259–295 (2017). https://doi.org/10.1016/j.ic.2016.10.011
8. Chatterjee, K., de Alfaro, L., Henzinger, T.A.: The Complexity of Stochastic Rabin and Streett Games. In: ICALP. pp. 878–890. Springer (2005). https://doi.org/10.1007/11523468_71
9. Chatterjee, K., Doyen, L., Randour, M., Raskin, J.F.: Looking at mean-payoff and total-payoff through windows. Information and Computation **242**, 25–52 (2015). https://doi.org/10.1016/j.ic.2015.03.010
10. Chatterjee, K., Henzinger, T.A., Horn, F.: Stochastic Games with Finitary Objectives. In: MFCS. pp. 34–54. Springer (2009). https://doi.org/10.1007/978-3-642-03816-7_4
11. Chatterjee, K., Piterman, N.: Combinations of Qualitative Winning for Stochastic Parity Games. In: CONCUR. LIPIcs, vol. 140, pp. 6:1–6:17 (2019). https://doi.org/10.4230/LIPIcs.CONCUR.2019.6
12. Doyen, L., Guha, S.: Algorithm and Strategy Construction for Sure-Almost-Sure Stochastic Parity Games. In: STACS. pp. 34:1–34:20 (2026). https://doi.org/10.4230/LIPICS.STACS.2026.34
13. Ehrenfeucht, A., Mycielski, J.: Positional Strategies for Mean Payoff Games. Int. Journal of Game Theory **8**(2), 109–113 (1979). https://doi.org/10.1007/BF01768705
14. Gaba, P., Guha, S.: Sure-almost-sure and Sure-limit-sure Window Mean Payoff in Markov Decision Processes (2026), https://arxiv.org/abs/2605.12191
15. Grädel, E., Thomas, W., Wilke, T. (eds.): Automata, Logics, and Infinite Games: A Guide to Current Research, Lecture Notes in Computer Science, vol. 2500. Springer (2002). https://doi.org/10.1007/3-540-36387-4
16. Zwick, U., Paterson, M.: The Complexity of Mean Payoff Games on Graphs. Theoretical Computer Science **158**(1&2), 343–359 (1996). https://doi.org/10.1016/0304-3975(95)00188-3