Multidimensional beyond worst-case and almost-sure problems for mean-payoff objectives

Lorenzo Clemente
Université Libre de Bruxelles
Brussels, Belgium

Jean-François Raskin
Université Libre de Bruxelles
Brussels, Belgium

Abstract—The *beyond worst-case threshold problem (BWC)*, recently introduced by Bruyère *et al.*, asks given a quantitative game graph for the synthesis of a strategy that i) enforces some minimal level of performance against any adversary, and ii) achieves a good expectation against a stochastic model of the adversary. They solved the BWC problem for finite-memory strategies and unidimensional mean-payoff objectives and they showed membership of the problem in NP∩coNP. They also noted that infinite-memory strategies are more powerful than finite-memory ones, but the respective threshold problem was left open.

We extend these results in several directions. First, we consider multidimensional mean-payoff objectives. Second, we study both finite-memory and infinite-memory strategies. We show that the multidimensional BWC problem is coNP-complete in both cases. Third, in the special case when the worst-case objective is unidimensional (but the expectation objective is still multidimensional) we show that the complexity decreases to NP∩coNP. This solves the infinite-memory threshold problem left open by Bruyère *et al.*, and this complexity cannot be improved without improving the currently known complexity of classical mean-payoff games. Finally, we introduce a natural relaxation of the BWC problem, the *beyond almost-sure threshold problem (BAS)*, which asks for the synthesis of a strategy that ensures some minimal level of performance with probability one and a good expectation against the stochastic model of the adversary. We show that the multidimensional BAS threshold problem is solvable in P.

I. INTRODUCTION

In a two-player mean-payoff game played on a weighted graph [1], [2], given a threshold $v \in \mathbb{Q}$, we must decide if there exists a strategy for Player 1 (the controller) to force plays with mean-payoff values larger than v, against any strategy of Player 2 (the environment). In the *beyond worst-case threshold problem (BWC)*, recently introduced by Bruyère *et al.* in [3], we are additionally given a stochastic model for the *nominal*, i.e. expected, behaviour of Player 2. Then we are asked, given two threshold values $\mu, \nu \in \mathbb{Q}$, to decide if there exists a strategy for Player 1 that forces (i) plays with a mean-payoff value larger than μ against any strategy of Player 2, and (ii) an expected mean-payoff value larger than ν when Player 2 plays according to the stochastic model of his nominal behaviour. In the BWC problem, we thus need to solve simultaneously a two player zero-sum game for the worst-case and an optimization problem where the adversary has been replaced by a stochastic model of his behaviour.

BWC is a natural problem: in practice, we want to build systems that ensure good performances when the environ-ment exhibits his nominal behaviour, and at the same time, that ensure some minimal performances no matter how the environment behaves. In [3], the BWC problem is solved for *finite-memory* strategies and *unidimensional* mean-payoff objectives, and shown to be in NP∩coNP. Also, it is noted there that *infinite-memory strategies* are more powerful than finite-memory ones, and that cannot even be approximated by the latter (already in the unidimensional case; cf. [4, Fig. 6] for an example). The respective threshold problem was left unsolved. We extend here these results in several directions.

A. Contributions

Our contributions are as follows. First, we consider *d-dimensional* mean-payoff objectives. Multiple dimensions are useful to model systems with *multiple objectives* that are potentially conflicting, and to analyze the possible trade-offs. For example, we may want to synthesize strategies that ensure a good QoS while keeping the energy consumption as low as possible. This extends the BWC problem with one additional level of conflicting trade-offs, which makes the analysis sub-stantially harder. Second, we study both finite-memory and infinite-memory strategies. We show that the multidimensional BWC problem is coNP-complete in both cases, and so not more expensive than the plain multidimensional mean-payoff games. This is obtained as a coNP reduction to the solu-tion of a linear system of inequalities of polynomial size. Correctness follows from non-trivial approximations results for finite/infinite-memory strategies inside end-components[1]. While in the unidimensional case optimal values for the expectation can always be achieved precisely (already by mem-oryless strategies), in our multidimensional setting this is not true anymore. To overcome this difficulty, we are able to show that achievable vectors can be approximated with arbitrary precision, which is sufficient for our analysis. Third, in the special case when the worst-case objective is unidimensional (but the expectation is still multidimensional), we show that the complexity decreases to NP∩coNP. This solves with optimal complexity the infinite-memory threshold problem left open in [3]. Finally, we introduce the *beyond almost-sure threshold problem (BAS)* which is a natural relaxation of the BWC problem. The BAS problem asks, given two threshold values

---

[1]Sub-MDPs which are strongly connected and closed w.r.t. the stochastic transitions.

$\vec{\mu}, \vec{\nu} \in \mathbb{Q}^d$, for the synthesis of a strategy for Player 1 that (i) ensures a mean-payoff larger than $\vec{\mu}$ almost surely, i.e. *with probability one*, and (ii) an expectation larger than $\vec{\nu}$ against the nominal behaviour of the environment. This problem has been independently considered (among other generalizations thereof) in [5]. We show that the multidimensional BAS threshold problem is solvable in P. As in the BWC problem, we reduce to a linear system of inequalities of polynomial size, but this time the reduction can be done in P.

B. Related works

Solutions to the expected unidimensional mean-payoff problem in Markov Decision Processes (MDP) can be found for example in [6], it can be solved in P, and pure memoryless strategies are sufficient to play optimally. The threshold problem for unidimensional mean-payoff games was first studied in [2], pure memoryless optimal strategies exist for both players, and the associated decision problem can be solved in NP∩coNP. As said above, BWC was introduced in [3] but studied only for finite memory strategies and unidimensional payoffs, the decision problem can be solve in NP∩coNP.

Multidimensional mean-payoff games are investigated in [7], [8], where it is shown that infinite-memory controllers are more powerful than finite-memory ones, and the finite-memory and general threshold problems are both coNP-complete. The expectation problem for multidimensional mean-payoff MDPs is in P, and finite-memory controllers always suffice [9]. Moreover, a recent study showed that one can add additional quantitative probability requirements for the mean-payoff to be above a certain threshold (while still optimizing the expectation), and that the resulting decision problem is P for the so-called *joint interpretation* (where the probability threshold is the same for all dimensions), and exponential for the *conjunction interpretation* (each dimension has a different probability threshold) [5] (cf. also [10]). In both cases, infinite-memory strategies are required to achieve the desired performance. Here, we study the *multidimensional mean-payoff BWC threshold problem*, for both finite-memory and arbitrary controllers. Our BWC threshold problem generalizes both the synthesis problem for multidimensional mean-payoff games and for multidimensional mean-payoff MDPs with no additional cost in worst-case computational complexity.

C. Illustrating example

Consider the following task system [11]: There are two configurations (0 and 1), and at each interaction between the controller and its environment, one new instance of two kind of tasks can be generated (0 and 1). The two tasks are generated with equal probability $1/2$ in the nominal behavior of the environment. Before serving pending task $k \in \{0, 1\}$, the system may decide to go from configuration i to configuration j at cost a_{ij}, for $i, j \in \{0, 1\}$, and then it has to serve the pending task k from the new configuration j at cost b_{jk}. Thus, the total cost is $a_{ij} + b_{jk}$. Costs are bidimensional: Each cost specifies an amount of time and energy; the actual parameters are shown in Fig. 1a. For example, from configuration 0,

task 0 takes 30 time units to complete and it consumes 2 energy units, while from the other configuration the same task takes 2 time units and 10 energy units. We are interested in synthesizing controllers that optimize the expected/worst-case mean (i.e., per task) time and energy. There are trade-offs between the two measures: If the controller decides to serve a task quickly then the system consumes a large amount of energy, and vice versa. To analyze this example, we rephrase it as the multidimensional mean-payoff MDP depicted in Fig. 1b. For example, state 0 represents the fact that the system is in configuration 0 waiting for a task to arrive, while in state $(0, 0)$ a task of the first type has arrived, and the controller needs to decide whether to serve it from the same configuration, or go to configuration 1. The objective of the controller is to guarantee worst-case mean time 24 under all circumstances, in that case the probabilities in the MDP are ignored and the probabilistic choice is replaced by an adversarial choice (we have thus a two-player zero sum game). Additionally, with the same strategy for the controller, we want to minimize the expected mean energy consumption in the nominal behaviour of the controller given by the stochastic model. If the controller decides to always serve tasks from configuration 0, then it ensures an expected mean energy consumption of 3, but under this strategy, the worst-case mean time is 60, which does not meet our worst-case objective of 24. A strategy for the controller that is good both for the worst-case and for the expectation can be obtained as follows: For two parameters $\alpha, \beta \in \mathbb{N}$, stay in configuration 0 for α consecutive tasks, then move to configuration 1 for β tasks, and then repeat. This ensures worst-case time $\frac{\alpha-1}{\alpha+\beta}60 + \frac{1}{\alpha+\beta}64 + \frac{\beta-1}{\alpha+\beta}8 + \frac{1}{\alpha+\beta}16$ and expected energy $\frac{\alpha-1}{\alpha+\beta}(\frac{1}{2}2 + \frac{1}{2}4) + \frac{1}{\alpha+\beta}(\frac{1}{2}4 + \frac{1}{2}6) + \frac{\beta-1}{\alpha+\beta}(\frac{1}{2}10 + \frac{1}{2}20) + \frac{1}{\alpha+\beta}(\frac{1}{2}16 + \frac{1}{2}26)$. By taking $\alpha = 1$ and $\beta = 3$, we obtain worst-case time 24 (thus meeting the requirement) and expected energy 14. Note the trade-off: To ensure a stronger guarantee on the mean time, we had to sacrifice the expected mean energy.
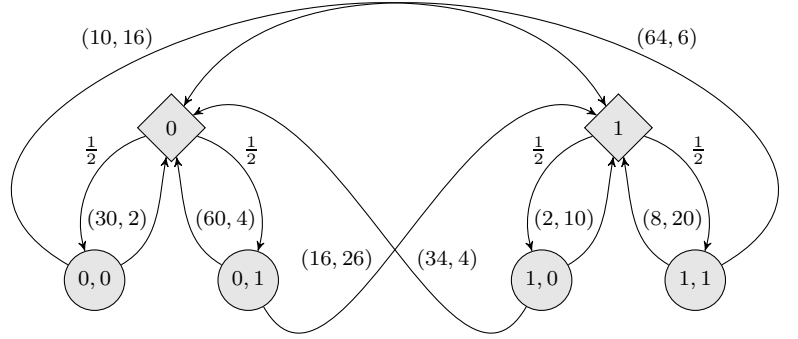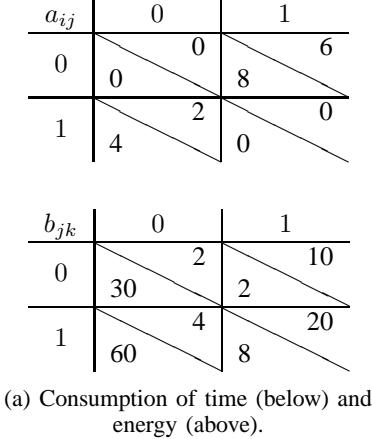
In this paper we address the problem of deciding the existence of controllers ensuring a worst-case (or almost-sure) threshold, while, at the same time, achieving a usually better expectation threshold under the nominal behavior of the environment. We consider the class of multidimensional mean-payoff objectives.

D. Structure of the paper

In Sec. II, we present the preliminaries that are necessary to define the BWC and BAS problems. In Sec. III, we solve the BWC problem both for finite and infinite memory strategies. In Sec. IV, we solve the BAS problem and show that finite memory strategies are sufficient to achieve the BAS threshold problem. Finally, in Sec.V we conclude with some final remarks. Full proofs can be found in the appendices A and B.

II. PRELIMINARIES

Let $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ be the set of natural, rational, and real numbers, respectively, and let $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{+\infty, -\infty\}$. For

a_{ij}	0	1
0	0 / 0	6 / 8
1	2 / 4	0 / 0

b_{jk}	0	1
0	2 / 30	10 / 2
1	4 / 60	20 / 8

(a) Consumption of time (below) and energy (above).

$(10, 16)$ $(64, 6)$
$\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$
$(30, 2)$ $(60, 4)$ $(2, 10)$ $(8, 20)$
$(16, 26)$ $(34, 4)$
$0, 0$ $0, 1$ $1, 0$ $1, 1$

(b) Example of multidimensional mean-payoff MDP.

Fig. 1: Illustrating example.

two vectors $\vec{\mu}$ and $\vec{\nu}$ of the same dimension and a comparison operator $\sim \in \{\leq, <, >, \geq\}$, we write $\vec{\mu} \sim \vec{\nu}$ for the component-wise application of $\sim$. In particular, $\vec{\mu} > \vec{0}$ means that *every component* of $\vec{\mu}$ is strictly positive. A *probability distribution* on A is a function $R : A \to \mathbb{Q}^{\geq 0}$ s.t. $\sum_{a \in A} R(a) = 1$. The *support* of R is $\mathsf{Supp}(R) = \{a \in A \mid R(a) > 0\}$. Let $\mathcal{D}(A)$ be the set of probability distributions on A.

A. Weighted graphs

A *multi-weighted graph* is a tuple $G = (d, S, E, w)$, where $d \geq 1$ is the dimension, S is a finite set of states, $E \subseteq S \times S$ is the set of directed edges, and $w : E \to \mathbb{Z}^d$ is a function assigning to each edge a weight vector. When $d = 1$, we refer to G just as a weighted graph. With $\vec{x}[i]$ we denote the i-th component of a vector $\vec{x}$. For a state $s \in S$, let $E(s) = \{t \mid (s, t) \in E\}$ be its set of *successors*. We assume that each state s has at least one successor. Let W be the largest absolute value of a weight appearing in the graph.

A *play* in G is an infinite sequence of states $\pi = s_0 s_1 \cdots$ s.t. $(s_i, s_{i+1}) \in E$ for every $i \geq 0$. Let $\Omega_{s_0}^\omega(G)$ be the set of plays in G starting at s_0, and let $\Omega^\omega(G)$ be the set of all plays of G. When G is clear from the context, we omit it. The *prefix* of length n of a play $\pi = s_0 s_1 \cdots$ is the finite sequence $\pi(n) = s_0 s_1 \cdots s_{n-1}$. For a set of states $T \subseteq S$, Let $\Omega^*(G, T)$ be the set of prefixes of plays in G ending in a state $s_{n-1} \in T$.

The *total payoff* and *mean payoff* up to length n of a play $\pi = s_0 s_1 \cdots$ (or prefix of length at least n) are defined as $\mathsf{TP}_n(\pi) = \sum_{i=0}^{n-1} w(s_i, s_{i+1})$ and $\mathsf{MP}_n(\pi) = \frac{1}{n} \mathsf{TP}_n(\pi)$, respectively. The (lim-inf) total and mean payoffs on an infinite play π are then defined as $\mathsf{TP}(\pi) := \liminf_{n \to \infty} \mathsf{TP}_n(\pi)$ and $\mathsf{MP}(\pi) := \liminf_{n \to \infty} \mathsf{MP}_n(\pi)$.

B. Markov decision processes

A *Markov decision process*, or MDP, is a tuple $\mathcal{G} = (G, S^C, S^R, R)$, where $G = (d, S, E, w)$ is a multi-weighted graph, $\{S^C, S^R\}$ is a partition of S into states belonging to either the Controller player or to the Random player, respectively, and $R : S^R \to \mathcal{D}(S)$ is a function assigning a distribution over S to states belonging to Random s.t., for every $s \in S^R$, $\mathsf{Supp}(R(s)) = E(s)$. We do not allow $R(s)$ to assign probability zero to any successor of s.[2] Let Q be the largest denominator used to represent probabilities in R. We use Q as a measure of complexity for representing R.

In order to discuss the complexity of strategies for Controller, we represent them as *stochastic Moore machines*. A strategy for a MDP $\mathcal{G} = (G, S^C, S^R, R)$ is a tuple $f = (M, \alpha, f_u, f_o)$, consisting of a set of memory states M, the initial memory distribution $\alpha \in \mathcal{D}(M)$, the stochastic memory update function $f_u : S \times M \to \mathcal{D}(M)$, and the stochastic output function $f_o : S^C \times M \to \mathcal{D}(S)$, where $\mathsf{Supp}(f_o(s, m)) \subseteq E(s)$ for every $s \in S^C$ and $m \in M$. The update function is extended to sequences $f_u^* : S^* \to \mathcal{D}(M)$ inductively as $f_u^*(\varepsilon) = \alpha$ and $f_u^*(\pi s)(m') = \sum_{m \in M} f_u^*(\pi)(m) f_u(s, m)(m')$. The output function on sequences $f_o^* : S^* S^C \to \mathcal{D}(S)$ is defined as $f_o^*(\pi s)(s') = \sum_{m \in M} f_u^*(\pi)(m) f_o(s, m)(s')$. A play $\pi = s_0 s_1 \cdots$ is *consistent* with a Controller's strategy f if, and only if, for every i s.t. $s_i \in S^C$, we have $s_{i+1} \in \mathsf{Supp}(f_o^*(s_0 s_1 \cdots s_i))$. Given a state s_0 and a Controller's strategy f, the set of *outcomes* $\Omega_{s_0}^f(G)$ is the set of plays starting at s_0 which are consistent with f.

A strategy f is *pure* iff $\mathsf{Supp}(f_u(s, m))$ and $\mathsf{Supp}(f_o(s, m))$ are both singletons. A strategy f is *memoryless* iff $|M| = 1$, *finite-memory* iff $|M| < \infty$, and *infinite-memory* iff $|M| = \infty$.

---

[2]This restriction will simplify the presentation. It is not present in [3], but it can be easily lifted.

Let $\Delta(\mathcal{G})$, $\Delta_P(\mathcal{G})$, $\Delta_F(\mathcal{G})$, and $\Delta_{PF}(\mathcal{G})$ be the sets of all, resp., pure, finite-memory, and pure finite-memory strategies.

C. Markov chains

A *Markov chain* is an MDP where no state belongs to Controller, i.e., $S^C = \emptyset$, and in this case we just write $\mathcal{G} = (G, R)$. An *event* is a measurable set of plays $A \subseteq \Omega^\omega(G)$. Given a state s_0 and an event $A \subseteq \Omega^\omega_{s_0}(G)$, let $\mathbb{P}^{\mathcal{G}}_{s_0}[A]$ be the probability that a play starting in s_0 belongs to A, which exists and it is unique by Carathéodory's extension theorem [12]. An even is *almost sure* if it has probability 1. For a measurable payoff function $v : \Omega^\omega(G) \to \mathbb{R}^d_{\pm\infty}$, let $\mathbb{E}^{\mathcal{G}}_{s_0}[v]$ be the expected value of v of a play starting in s_0.

A Markov chain $\mathcal{G}$ is *unichain* if it contains exactly one bottom strongly connected component (BSCC). Therefore, if $\mathcal{G}$ is unichain, then all states in its unique BSCC are visited infinitely often almost surely, and the mean payoff equals its expected value almost surely.

Given a MDP $\mathcal{G} = (G, S^C, S^R, R)$ and a strategy f for Controller represented as the stochastic Moore machine (M, α, f_u, f_o), let the *induced Markov chain* be $\mathcal{G}[f] = (G', R')$, where $G' = (d, S \times M, w', E')$ with $((s, m), (s', m')) \in E'$ iff $(s, s') \in E$, $m' \in \mathsf{Supp}(f_u(s, m))$, and $s' \in \mathsf{Supp}(f_o(s, m))$ whenever $s \in S^C$, $w'((s, m), (s', m')) = w(s, s')$ for every $((s, m), (s', m')) \in E'$, $R'(s, m)(s', m') = R(s)(s') \cdot f_u(s, m)(m')$ for every $s \in S^R$, and $R'(s, m)(s', m') = f_o(s, m)(s') \cdot f_u(s, m)(m')$ for every $s \in S^C$. Note that $\mathcal{G}[f]$ is finite iff f is finite-memory. By a slight abuse of terminology, we say that a strategy f is *unichain* if $\mathcal{G}[f]$ is unichain. Plays in $\mathcal{G}[f]$ can be mapped to plays in $\mathcal{G}$ by a projection operator $\mathsf{proj}(\cdot) : \Omega^\omega(G') \to \Omega^\omega(G)$ which discards the memory of f. Given a state s_0, a Controller's strategy f, and an event $A \subseteq \Omega^\omega_{s_0}$, let $\mathbb{P}^{\mathcal{G}}_{s_0, f}[A] := \mathbb{P}^{\mathcal{G}[f]}_{s_0}[\mathsf{proj}^{-1}(A)]$. For a measurable payoff function $v : \Omega^\omega(G) \to \mathbb{R}^d_{\pm\infty}$, let $\mathbb{E}^{\mathcal{G}}_{s_0, f}[v] := \mathbb{E}^{\mathcal{G}[f]}_{s_0}[v']$, where $v'(\pi) := v(\mathsf{proj}(\pi))$.

D. End-components

A *end-component* (EC) of a MDP $\mathcal{G}$ is a set of states $U \subseteq S$ s.t. a) the induced sub-graph $(U, E \cap U \times U)$ is strongly-connected, and b) for any stochastic state $s \in U \cap S^R$, $E(s) \subseteq U$. Thus, Controller can surely keep the game inside an EC, and almost surely visits all states therein. For an end-component U of $\mathcal{G}$, we denote by $\mathcal{G} \downharpoonright U$ the MDP obtained by restricting $\mathcal{G}$ to U in the natural way. ECs are central in the analysis of MDPs thanks to the following result.

Proposition 1 (cf. [13]). *For any Controller's strategy $f \in \Delta(\mathcal{G})$, the set of states visited infinitely often when playing according to f is almost surely an EC.*
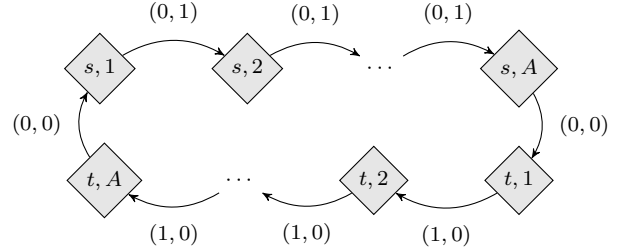
E. Expected-value objective

For a MDP $\mathcal{G}$, a starting state s_0, and Controller's strategy $f \in \Delta(\mathcal{G})$, the set of *expected-value achievable solutions for f* is $\mathsf{ExpSol}^+_{\mathcal{G}}(s_0, f) = \{\vec{\nu} \in \mathbb{R}^d \mid \mathbb{E}^{\mathcal{G}}_{s_0, f}[\mathsf{MP}] > \vec{\nu}\}$, i.e., it is the set of vectors $\vec{\nu}$ s.t. Controller can guarantee an
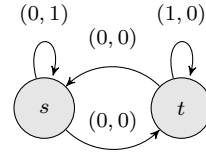
(a) A MDP reduced to one EC.
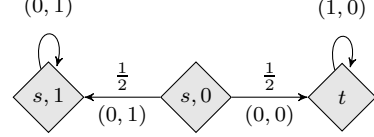
(b) Exact strategy inducing two BSCCs.

(c) Approximate finite-memory strategy inducing one BSCC.

Fig. 2: Approximating the expectation inside ECs.

expected mean payoff $> \vec{\nu}$ from state s_0 by playing f. The set of *expected-value achievable solutions* is $\mathsf{ExpSol}^+_{\mathcal{G}}(s_0) = \bigcup_{f \in \Delta(\mathcal{G})} \mathsf{ExpSol}^+_{\mathcal{G}}(s_0, f)$. Given a state s_0 and rational threshold vector $\vec{\nu} \in \mathbb{Q}^d$, the *expected-value threshold problem* asks whether $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G}}(s_0)$.

Theorem 1 ([9]). *The expected-value threshold problem for multidimensional mean-payoff MDPs is in* P.

While randomized finite-memory strategies are both necessary and sufficient in general for achieving a given expected mean payoff, in ECs we can use randomized finite-memory *unichain* strategies to *approximate* achievable vectors. Being unichain ensures that the mean payoff equals the expectation almost surely. By standard convergence results in Markov chains, this entails that by playing such a strategy for sufficiently long time we obtain an average mean payoff close to the expectation with high probability. We crucially use this property in the constructions leading to the main results of Sec. III and IV; cf. Lemmas 3, 5, and 8.

Example 1. We illustrate the idea in the single end-component MDP in Fig. 2a (cf. [8, Fig. 3]). There exists a simple randomized 2-memory strategy f achieving expected mean payoff precisely $(\frac{1}{2}, \frac{1}{2})$ which decides, with equal probability, whether to stay forever in s or in t. However, the induced Markov chain has two BSCCs; cf. Fig. 2b. While intuitively no pure finite-memory strategy can achieve mean payoff exactly equal $(\frac{1}{2}, \frac{1}{2})$ in this example, finite-memory unichain strategies

can approximate this value. For a parameter $A \in \mathbb{N}$, consider the strategy g_A which stays in s for A steps, and then goes to t, stays in t for A steps, and then goes back to s, and repeats this scheme forever. The induced Markov chain has only one BSCC, thus g_A is unichain; cf. Fig. 2c. The strategy g_A achieves expected (and worst-case) mean payoff $\left(\frac{A}{2A+2}, \frac{A}{2A+2}\right)$, which converges from below to $(\frac{1}{2}, \frac{1}{2})$ as $A \to \infty$.

Lemma 1. *Let $\mathcal{G}$ be a multidimensional mean-payoff MDP, let s_0 be a state in an EC U thereof, and let $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G} \restriction U}(s_0)$ be an expectation vector achievable by remaining inside U. There exists a finite-memory unichain strategy $g \in \Delta_F(\mathcal{G})$ achieving the same expectation $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G} \restriction U}(s_0, g)$.*

Remark 1. In the lemma above, we can take g to be even a *pure* finite-memory unichain strategy. This can be obtained by a de-randomization technique at the cost of introducing extra memory of size exponential in the number of the states controlled by the player. However, we do not need this stronger result in the rest of the paper, and we content ourselves with randomized strategies for simplicity.

Proof sketch. By the results of [9], there exists a randomized finite-memory strategy f achieving expected mean payoff $\vec{\nu}^* > \vec{\nu}$ which surely stays inside U. However, $(\mathcal{G} \restriction U)[f]$ is not unichain in general. By Proposition 1, the set of states visited infinitely often by a play in $(\mathcal{G} \restriction U)[f]$ is an EC almost surely. Since there are finitely many different ECs, there are probabilities $\alpha_1, \ldots, \alpha_n > 0$ and ECs $U_1, \ldots, U_n \subseteq U$ s.t. the set of states visited infinitely often by a play in $(\mathcal{G} \restriction U)[f]$ is U_1 with probability α_1, ..., U_n with probability α_n. By Proposition 1, $\alpha_1 + \cdots + \alpha_n = 1$. In the first step, we define a "local" randomized memoryless strategy g_i which plays as f once inside U_i. No approximation is introduced in this step. In the second step, we combine the local randomized memoryless strategies g_i's above. We build a randomized finite-memory strategy g which cycles between $U_1, \ldots, U_n$ and plays according to g_i inside each U_i a fraction $\approx \alpha_i$ of the time. This is possible since U_i, U_j are almost surely mutually inter-reachable due to the fact that we are always inside the EC U. By construction, $(\mathcal{G} \restriction U)[g]$ is unichain since g cycles between all the ECs $U_1, \ldots, U_n$. Moreover, for every $\varepsilon > 0$, we can make the expected fraction of time spent changing component smaller than ε. Thus, g achieves expected mean payoff at least $(1 - \varepsilon) \cdot \vec{\nu}^* - (W, \ldots, W) \cdot \varepsilon$, where W is the largest absolute value of any weight in $\mathcal{G}$. The latter quantity can be made $> \vec{\nu}$ for sufficiently small $\varepsilon > 0$. $\square$

F. Worst-case objective

For a MDP $\mathcal{G}$, a starting state s_0, and a Controller's strategy $f \in \Delta(\mathcal{G})$, the set of *worst-case achievable solutions for f* is defined as $\mathsf{WCSol}^+_{\mathcal{G}}(s_0, f) = \{\vec{\mu} \in \mathbb{R}^d \mid \forall \pi \in \Omega^f_{s_0} \cdot \mathsf{MP}(\pi) > \vec{\mu}\}$, i.e., it is the set of vectors $\vec{\mu}$ s.t. Controller can surely guarantee a mean payoff $> \vec{\mu}$ from state s_0 by playing f. The set of *worst-case achievable solutions* is $\mathsf{WCSol}^+_{\mathcal{G}}(s_0) = \bigcup_{f \in \Delta(\mathcal{G})} \mathsf{WCSol}^+_{\mathcal{G}}(s_0, f)$. Given a

state s_0 and rational threshold vector $\vec{\mu} \in \mathbb{Q}^d$, the *worst-case threshold problem* asks whether $\vec{\mu} \in \mathsf{WCSol}^+_{\mathcal{G}}(s_0)$.

With this worst-case interpretation, the randomized choices in the MDP are replaced by purely adversarial ones, and the MDP can thus be viewed as a two-player zero-sum game. While infinite-memory strategies are more powerful than finite-memory ones for the worst-case objective, the latter suffice to approximate achievable vectors. We make extensive use of this property in Sec. III-A where we restrict our attention to finite-memory strategies.

Lemma 2 (cf. Lemma 15 of [8])**.** *Let $\mathcal{G}$ be a multidimensional mean-payoff MDP, s_0 a state therein, and let $\vec{\mu} \in \mathsf{WCSol}^+_{\mathcal{G}}(s_0)$. There exists a pure finite-memory strategy $f \in \Delta_{PF}(\mathcal{G})$ for Controller s.t. $\vec{\mu} \in \mathsf{WCSol}^+_{\mathcal{G}}(s_0, f)$.*

The finite-memory strategy threshold problem for multidimensional mean-payoff games is coNP-complete [7], [8]. By the lemma above, finite memory controllers suffice in our setting, and we obtain the following complexity characterization.

Theorem 2 ([7], [8])**.** *The worst-case threshold problem for multidimensional mean-payoff MDPs is* coNP*-complete.*

In the unidimensional case, memoryless strategies suffice for both players [1], [2], and the complexity is NP∩coNP (and even UP∩coUP [11], [14]). It is open since long time whether this problem is in P.

Theorem 3 ([1], [2])**.** *The worst-case threshold problem for unidimensional mean-payoff MDPs is in* NP∩coNP*.*

III. BEYOND WORST-CASE SYNTHESIS

We generalize [3] to the multidimensional setting. Given a MDP $\mathcal{G}$, a starting state s_0, and a Controller's strategy f, the set of *beyond worst-case achievable solutions for f*, denoted $\mathsf{BWCSol}^+_{\mathcal{G}}(s_0, f)$, is the set of pairs of vectors $(\vec{\mu}; \vec{\nu}) \in \mathbb{R}^{2d}$ s.t. f surely guarantees a worst-case mean payoff $> \vec{\mu}$, and achieves an expected mean payoff $> \vec{\nu}$ starting from s_0,

$$\mathsf{BWCSol}^+_{\mathcal{G}}(s_0, f) = \left\{ (\vec{\mu}; \vec{\nu}) \in \mathbb{R}^{2d} \;\middle|\; \begin{array}{c} \vec{\mu} \in \mathsf{WCSol}^+_{\mathcal{G}}(s_0, f) \\ \text{and} \\ \vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G}}(s_0, f) \end{array} \right\}$$

Let $\mathsf{BWCSol}^+_{\mathcal{G}}(s_0) = \bigcup_{f \in \Delta(\mathcal{G})} \mathsf{BWCSol}^+_{\mathcal{G}}(s_0, f)$ be the set of *beyond worst-case achievable solutions*. Given a starting state s_0 and a pair of threshold vectors $(\vec{\mu}; \vec{\nu}) \in \mathbb{R}^{2d}$, the *beyond worst-case threshold problem* (BWC) asks whether $(\vec{\mu}; \vec{\nu}) \in \mathsf{BWCSol}^+_{\mathcal{G}}(s_0)$.

Remark 2. We assume w.l.o.g. that $\vec{\mu} = \vec{0}$. This follows by shifting each component by an appropriate amount. We further assume w.l.o.g. that $\vec{\nu} \geq \vec{0}$. This follows from the fact that, since the mean payoff is surely $> \vec{0}$ by the worst-case objective, then also the expectation is $> \vec{0}$.

Remark 3. We say that $\mathcal{G}$ is *pruned* if $\vec{0} \in \mathsf{BWCSol}^+_{\mathcal{G}}(s)$ for every state s therein. Controller cannot satisfy the BWC objective if she ever visits a state s not satisfying the worst-case objective. Many of our results are thus stated under the condition that $\mathcal{G}$ is pruned. However, pruning an MDP, i.e.,
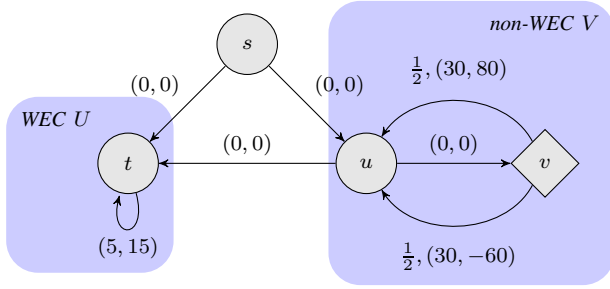
Fig. 3: Running example.

removing those states which are losing w.r.t. the worst-case objective, requires solving a mean-payoff game, and this will have a crucial impact on the complexity.

The finite-memory threshold problem for the unidimensional beyond worst-case problem has been studied in [3].

Theorem 4 ([3])**.** *The finite-memory threshold problem for the unidimensional beyond worst-case problem for mean-payoff objectives is in* NP∩coNP.

A. Finite-memory synthesis

In this section, we address the problem of deciding whether there exists a finite-memory strategy for the BWC problem in the multidimensional setting. By Proposition 1, we know that the set of states visited infinitely often by any strategy (not necessarily a finite-memory one) is almost surely an EC. The crucial observation is that, when restricted to finite-memory, the same holds for ECs of a special kind. An EC U is *winning* (WEC) iff Controller can surely guarantee the worst-case threshold $> \vec{0}$ when constrained to remain in U, starting from any state therein. Whether a EC is winning depends on the worst-case objective alone.

The following proposition is central in the analysis of the BWC problem for finite-memory strategies; cf. [3, Lemma 4] in the unidimensional case.

Proposition 2. *Let f be a* finite-memory *strategy satisfying the worst-case threshold problem. The set of states visited infinitely often under f is almost surely a* winning *EC.*

Running example. As a simple example that will be used through the rest of the paper, consider the MDP in Fig. 3. There are only two ECs U and V, of which U is winning, but V is not. Indeed, from v the adversary can always select the lower edge with payoff $(30, -60)$. In U we can achieve expectation $(5, 15)$, and from V we can achieve expectation $(15, 5)$. Therefore, according to the lemma above, any finite-memory strategy satisfying the worst-case objective will eventually go to U almost surely.

We proceed by analyzing WECs separately in Sec. III-A1, and then we tackle general MPDs in Sec. III-A2. This will yield our complexity result in Sec. III-A3.

1) Inside a WEC: We show that inside WECs finite-memory strategies always suffice for the BWC objective. In particular, the threshold problem in WECs immediately reduces to an expectation threshold problem.

Lemma 3. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP, let s_0 be a state in a WEC W of $\mathcal{G}$, and let $\vec{\nu} \in$ ExpSol$_{\mathcal{G}\downarrow W}^{+}(s_0)$ with $\vec{\nu} \geq \vec{0}$ be an expectation achievable by remaining inside W. There exists a randomized finite-memory strategy $h \in \Delta_F(\mathcal{G})$ s.t. $(\vec{0}; \vec{\nu}) \in$ BWCSol$_{\mathcal{G}\downarrow W}^{+}(s_0, h)$ that also remains inside W.*

Remark 4. The statement of the lemma holds even with h a *pure* finite-memory strategy, by applying Remark 1 when constructing the expectation strategy which is part of h. However, randomized strategies suffice for our purposes.

We use finite-memory strategies defined in WECs (such as h above) when constructing a global BWC strategy in the analysis of arbitrary MDPs in Sec. III-A2. The construction of h is done in a way analogous to the proof of Theorem 5 in [3]; cf. App. B for the details. However, the analysis in the multidimensional case is considerably more difficult than in previous work. It crucially relies on Lemma 1 for the extraction of *finite-memory unichain* strategies approximating the expectation objective inside ECs. Note that in the unidimensional case of [3] optimal expectation values can be reached exactly already by *pure memoryless unichain strategies* (no approximation needed). This is an key technical difference between our multidimensional setting and the unidimensional one of [3].

2) General case: We reduce the finite-memory BWC problem to the solution of a system of linear inequalities. This is similar to the solution of the multidimensional expectation problem presented in [9]. When only the expectation is considered, the intuition is that a "global expectation" is obtained by combining together "local expectations" achieved in ECs. Thus, a strategy for the expectation works in two phases:

Phase I: Reach ECs with appropriate probabilities.
Phase II: Once inside an EC, switch to a local expectation strategy to achieve the right "local expectation".

In the BWC problem, we need to enforce two extra conditions: First, only "local expectations" from *winning* ECs should be considered (by Proposition 2 finite-memory controllers cannot stay in a non-WEC forever with non-zero probability). Second, "local expectations" should be $> \vec{0}$ in order to satisfy the worst-case objective (a negative "local expectation" would violate the worst-case objective). Accordingly, a strategy for the BWC problem behaves as follows:

Phase I: Reach WECs with appropriate probabilities.
Phase II: Once inside a WEC, switch to a local BWC strategy to achieve the right "local expectation" $> \vec{0}$.

We write a system of linear inequalities expressing this two-phase decomposition. W.l.o.g. we assume that state s_0 belongs to Controller, and that all WECs are reachable with positive probability from s_0 (unreachable states can be removed). Consider the system T in Fig. 4. For each state $s \in S$ we have a variable y_s, and for each edge $(s, t) \in E$ we have

$$1_{s_0}(s) + \sum_{(r,s)\in E} y_{rs} = \sum_{(s,t)\in E} y_{st} + y_s \qquad \forall s \in S \qquad \text{(A1)}$$

$$y_{st} = R(s)(t) \cdot \Big(\sum_{(r,s)\in E} y_{rs} - y_s \Big) \qquad \forall (s,t) \in E \text{ with } s \in S^R \qquad \text{(A1')}$$

$$\sum_{\text{MWEC } U} \sum_{s\in U} y_s = 1 \qquad \text{(A2)}$$

$$\sum_{s\in U} y_s = \sum_{(r,s)\in E\cap U\times U} x_{rs} \qquad \forall \text{ MWEC } U \qquad \text{(B)}$$

$$\sum_{(r,s)\in E} x_{rs} = \sum_{(s,t)\in E} x_{st} \qquad \forall s \in S \qquad \text{(C1)}$$

$$x_{st} = R(s)(t) \cdot \sum_{(r,s)\in E} x_{rs} \qquad \forall (s,t) \in E \text{ with } s \in S^R \qquad \text{(C1')}$$

$$\sum_{(s,t)\in E} x_{st} \cdot w(s,t)[i] > \vec{\nu}[i] \qquad \forall (1 \le i \le n) \qquad \text{(C2)}$$

$$\sum_{(s,t)\in E\cap U\times U} x_{st} \cdot w(s,t)[i] > 0 \qquad \forall \text{ MWEC } U, 1 \le i \le n \qquad \text{(C3)}$$

Fig. 4: Linear system T for the BWC finite-memory threshold problem.

variables x_{st} and y_{st}. System T can be divided into three parts. The first part consists of Equations (A1)–(A2). Variable y_s represents the probability that, upon visiting state s, we switch to Phase II. Variables y_{st}'s are used to express flow conditions. In Eq. (A1) we put an initial flow of 1 in s_0, and we require that the total incoming flow to a state equals the outgoing flow (including the leak y_s). In Eq. (A1') ensures that the outgoing flow through an edge y_{st} from a stochastic state s is a fixed fraction of the incoming flow. Finally, Eq. (A2) states that we switch to Phase II in a WEC almost surely.

Before explaining the other two parts of T, we need to introduce maximal WECs. A *maximal WECs* (MWEC) is a WEC which is not strictly included into another WEC. The restriction to MWECs is crucial for complexity. The second part of T consists of Eq. (B) and it provides a link between Phase I and Phase II. Variable x_{st} represents the long-run frequency of edge (s,t). Eq. (B) links the transient behaviour before switching inside a certain MWEC and the steady state behaviour once inside it. More precisely, it guarantees that the probability to switch inside a certain MWEC equals the total long-run frequency of all edges in the MWEC.

Finally, the remaining equations make up the third part of T. Eq. (C1) is a flow condition for the x_{st}'s, stating that the incoming flow to a state equals the outgoing flow. Eq. (C1') forces the flow to respect the probabilities of stochastic states. Eq. (C2) guarantees that the expected mean payoff is $> \vec{\nu}$, as required. Eq. (C3) needs some justification. It is specific to our setting and it does not follow from [9]. This equation specifies that the expected mean payoff is $> \vec{0}$ *inside every MWECs*. We need to ensure that only "local" expected mean payoffs $> \vec{0}$ should be considered in WECs, in order to be able to apply the results from the previous Sec. III-A1. Eq. (C3) imposes

a seemingly strong constraint by requiring that *all WECs* are visited infinitely often with positive probability. Ideally, we would like to guess which are the MWECs which need to be visited infinitely often with positive probability, but this would not yield a good complexity, since there are exponentially many different sets of MWECs. Instead, we require that *every* MWEC is visited infinitely often with some positive probability. Since we are only interested in approximating the expectation, it is always possible to put an arbitrary small total probability on MWECs that do not contribute to the "global" mean payoff. This is formalized below.

Proposition 3. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP. If there exists a finite-memory strategy h s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h)$, then there exists a finite-memory strategy h^* with the same property, and such that, for every MWEC U, the set of states visited infinitely often by h^* is a subset of U with positive probability.*

Proof. Since by assumption all MWEC are reachable with positive probability from s_0, for every MWEC U there exists a strategy f_U reaching U with positive probability from s_0. Moreover, since U is a WEC, there exists a strategy f_U^{wc} for the worst-case objective $> \vec{0}$ that surely remains in U. Let f^{wc} be a worst-case strategy winning everywhere (it exists since $\mathcal{G}$ is pruned by assumption). We construct the following strategy f_N parametrized by a natural number $N > 0$:

- Choose a MWEC U uniformly at random.
- Play f_U for N steps.
  - If after N steps the play is in U, then switch to f_U^{wc}.
  - Otherwise, switch to f^{wc}.

By construction f_N is winning for the worst-case for every $N > 0$. Moreover, it is easy to see that there exists an

N^* sufficiently large s.t., for every MWEC U, f_{N^*} visits U infinitely often with positive probability.

Finally, the strategy h^* plays with probability $p > 0$ according to f_{N^*}, and otherwise according to h. Since both f_{N^*} and h are winning for the worst-case, so it is h^*. The expected mean payoff of h^* converges from below to the expected mean payoff of h for $p > 0$ sufficiently small. Therefore, there exists $p > 0$ s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h^*)$. $\qquad\square$

We now state the correctness of the reduction.

Lemma 4. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP, let $s_0 \in S$, and let $\vec{\nu} \geq \vec{0}$. There exists a finite-memory strategy h s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h)$, if, and only if, the system T has a non-negative solution.*

The rest of this section is devoted to the proof of the lemma above. Both directions are non-trivial. For the right-to-left direction, we need to explain which kind of strategies can be extracted from a non-negative solution of T. The following lemma shows that from a non-negative solution of T we can extract a strategy for the expectation combining only "local mean payoffs" $> \vec{0}$ and visiting infinitely often each MWEC with positive probability.

Proposition 4. *If T has a non-negative solution, then there exists a finite-memory strategy $\hat{h}$ s.t. $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}^+(s_0, \hat{h})$, and*

1) *For every MWEC U, there is a probability $y_U^* > 0$ s.t. the set of states visited infinitely often by $\hat{h}$ is a subset of U with probability y_U^*.*
2) *Once $\hat{h}$ reaches the MWEC U, it achieves expected mean payoff $\vec{\nu}_U > \vec{0}$.*
3) *$\sum_{\text{MWEC } U} y_U^* \cdot \vec{\nu}_U > \vec{\nu}$.*

Proof. Let $\{y_s^*\}_{s \in S}$, $\{y_{st}^*\}_{(s,t) \in E}$, and $\{x_{st}^*\}_{(s,t) \in E}$ be a non-negative solution to T. Proposition 4.2 of [9] essentially shows how to construct from the solution above a finite-memory strategy $\hat{h}$ s.t. $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}^+(s_0, \hat{h})$.

For a MWEC U, let

$$y_U^* = \sum_{s \in U} y_s^* \tag{1}$$

By Eq. (C3), for every MWEC U there exist $s, t \in U$ s.t. $x_{st}^* > 0$. Together with Eq. (B), this implies that $y_U^* > 0$, which proves Point 1.

For a MWEC U, let

$$\vec{\nu}_U = \sum_{(s,t) \in E \cap U \times U} x_{st}^* \cdot w(s,t) \tag{2}$$

and notice that $\vec{\nu}_U$ is the expected mean payoff of $\hat{h}$ once inside U. By Eq. (C3), $\vec{\nu}_U > \vec{0}$, which proves Point 2.

Eq. (B) implies that $\hat{h}$ eventually stays forever inside a WEC almost surely. Consequently, $\sum_{\text{MWEC } U} y_U^* = 1$. Since states visited infinitely often with probability zero do not contribute to the expected mean payoff, it suffices to look at MWECs. By the prefix independence of the mean payoff value function, and since MWEC U is reached with probability y_U^*, strategy

$\hat{h}$ achieves expected mean payoff $\sum_{\text{MWEC} U} y_U^* \cdot \vec{\nu}_U$. By Point 1), the latter quantity is $> \vec{\nu}$. $\qquad\square$

We are now ready to prove Lemma 4.

Proof of Lemma 4. For the left-to-right direction, assume that h is a finite-memory strategy guaranteeing $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s, h)$. Proposition 4.4 of [9] essentially shows that any strategy satisfying the expectation objective $> \vec{\nu}$ induces a solution to T satisfying Equations (A1)–(C2), except that Eq. (B) should be interpreted over MECs (instead of MWECs). (This follows from the fact that the set of states visited infinitely often by any strategy is an EC almost surely; cf. Proposition 1.) However, since $\vec{0} \in \mathsf{WCSol}_{\mathcal{G}}^+(s, h)$ and h is finite-memory, we can apply Proposition 2 and deduce that h visits infinitely often a *winning* EC almost surely. Thus Eq. (B) is satisfied even over MWECs.

It remains to address Eq. (C3). By Proposition 3, there exists a strategy h^* s.t., for every MWEC U, h^* eventually stays forever in U with a positive probability. This implies that, when constructing a solution to T induced by h^* (as above), for every MWEC U and $s, t \in U$, $x_{st}^* > 0$. Moreover, since h^* is winning for the worst-case, it achieves an expected mean payoff $> \vec{0}$ in U, and thus Eq. (C3) is satisfied.

For the right-to-left direction, assume that T has a non-negative solution. Let $\hat{h}$ be the strategy in $\mathcal{G}$ given by Proposition 4. For every MWEC U, let y_U^* and $\vec{\nu}_U$ be as given in the statement of the proposition. While $\hat{h}$ alone is not sufficient to show $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s)$ since it does not satisfy the worst-case objective in general, we show how to construct from it another finite-memory strategy h^{cmb} ensuring the BWC objective. The latter strategy is obtained by combining together the following strategies:

- Let h^{wc} be a finite-memory strategy in $\mathcal{G}$ ensuring the worst-case mean payoff $\vec{0} \in \mathsf{WCSol}_{\mathcal{G}}^+(t, h^{wc})$ from every state t in $\mathcal{G}$. This is possible since $\mathcal{G}$ is pruned.
- For each MWEC U, let h_U be a finite-memory strategy s.t. $(\vec{0}; \vec{\nu}_U) \in \mathsf{BWCSol}_{\mathcal{G}}^+(t, h_U)$ for every state $t \in U$. This strategy can be obtained as follows. Let $\mathcal{G} \downarrow U$ be the game $\mathcal{G}$ restricted to the EC U. By Point 2 of Proposition 4, $\vec{\nu}_U \in \mathsf{ExpSol}_{\mathcal{G} \downarrow U}^+(t_0, h_U)$ for some state $t_0 \in U$. Since U is an EC, $\vec{\nu}_U \in \mathsf{ExpSol}_{\mathcal{G} \downarrow U}^+(t, h_U)$ for *every* state $t \in U$. Since $\vec{\nu}_U > \vec{0}$, we can apply Lemma 3 for every $t \in U$, and obtain a strategy h_t s.t. $(\vec{0}; \vec{\nu}_U) \in \mathsf{BWCSol}_{\mathcal{G} \downarrow U}^+(t, h_t)$. Let h_U be the finite-memory strategy in $\mathcal{G} \downarrow U$ that plays according to h_t when starting from state t. Clearly, $(\vec{0}; \vec{\nu}_U) \in \mathsf{BWCSol}_{\mathcal{G} \downarrow U}^+(t, h_U)$.

Consider the strategy h_N^{cmb} parameterized by a natural number $N > 0$ which is defined as follows:

1) Play according to $\hat{h}$ for N steps.
2) After N steps:
   2a) If we are inside the MWEC U, then switch to h_U.
   2b) Otherwise, play according to h^{wc}.

We argue that h_N^{cmb} satisfies the beyond worst-case objective $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h_N^{cmb})$ for N large enough. For every N, h_N^{cmb} clearly satisfies the worst-case objective, since after

N steps it switches to a strategy that satisfies it by construction (by prefix-independence of the mean payoff objective). We now consider the expectation objective. By Point 1 of Proposition 4, the set of states visited infinitely often by $\hat{h}$ is a subset of the MWEC U with probability y_U^*. By taking N large enough, we can guarantee being inside U with probability arbitrarily close to y_U^*. By construction, h_U can be chosen to achieve expected mean payoff arbitrarily close to $\vec{\nu}_U$. Since h_N^{cmb} switches to h_U with probability arbitrarily close to y_U^*, h_N^{cmb} achieves expected mean payoff arbitrarily close to $\sum_{\text{MWEC } U} y_U^* \cdot \vec{\nu}_U$. By Point 3 of Proposition 4, the latter quantity is $> \vec{\nu}$. There exists N^* large enough s.t. $h_{N^*}^{cmb}$ achieves expected mean payoff $> \vec{\nu}$. Take $h^{cmb} = h_{N^*}^{cmb}$. As required, $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}^+(s_0, h^{cmb})$. $\square$

Running example. Since $U = \{t\}$ is a MWEC, while $V = \{u, v\}$ is not, finite memory strategies must go to U. Therefore, with finite memory we can ensure BWC threshold $((0,0);(0,9))$, but not $((0,0);(9,9))$ for example.

3) Complexity: We obtain the following complexity characterization for the threshold problem with finite-memory controllers.

Theorem 5. *The finite-memory multidimensional mean-payoff BWC threshold problem is* coNP-*complete.*

Proof. Pruning states where the worst-case objective cannot be satisfied requires solving multidimensional mean-payoff games, which can be done in coNP by Theorem 2. It has been already shown in [3] how the decomposition in MWEC can be performed in Pwith an oracle for solving mean-payoff games. Thus, the MWEC decomposition can be performed in coNP. System T has size polynomial in $\mathcal{G}$ (there are only polynomially many MWECs) and it can thus be produced in coNP. By Lemma 4, it suffices to solve system T, which can be done in polynomial time by linear programming. The lower bound follows directly from the fact that the multidimensional BWC threshold problem contains the worst-case as a subproblem; the latter is coNP-hard as recalled in Theorem 2. $\square$

The complexity of the BWC problem is dominated by the worst-case subproblem. We obtain an improved complexity by restricting the worst-case to be essentially unidimensional. Formally, we say that a BWC threshold $(\vec{\mu}; \vec{\nu}) \in \mathbb{R}^{2d}$ has *trivial worst-case component* i, with $1 \leq i \leq d$, iff $\vec{\mu}[i] = -W$, where W is the maximal absolute value of any weight in $\mathcal{G}$. We say that $(\vec{\mu}; \vec{\nu})$ *is essentially worst-case unidimensional* iff it has at most one non-trivial worst-case component. We can ignore trivial components when solving a worst-case threshold problem. Thus, the worst-case problem for essentially unidimensional thresholds reduces to a simple unidimensional worst-case problem. As recalled in Theorem 3, the latter can be solved in NP∩coNP, thus yielding the following improved complexity for the BWC problem.

Corollary 1. *The finite-memory multidimensional mean-payoff BWC threshold problem w.r.t.* essentially worst-case unidimensional thresholds *is in* NP∩coNP.

Since the *unidimensional* BWC problem, i.e., where all weights are unidimensional, is in NP∩coNP (cf. Theorem 4), this results shows that we can add a multidimensional expectation objective to a unidimensional worst-case obligation without an extra price in complexity. In particular, we can model complex situations like the task system presented in Sec. I-C, where the worst-case and expectation mean payoffs are along independent dimensions.

B. Infinite-memory synthesis

Already in the unidimensional case, infinite-memory strategies are more powerful than finite-memory ones (cf. [4, Fig. 6]). This is a consequence of the fact that finite-memory strategies for the BWC objective ultimately remain inside WECs almost surely (cf. Proposition 2). On the other hand, infinite-memory strategies can benefit from payoffs achievable inside arbitrary ECs. In this section, we address the problem of deciding whether there exists a general strategy, i.e., not necessarily finite-memory one, for the multidimensional BWC problem. This was left as an open problem, already in the unidimensional case [3]. As in the previous section, we first analyze ECs, and then general MDPs.

1) Inside an EC: The lemma below is a direct generalization of Lemma 2 to arbitrary ECs. While for WECs we could construct finite-memory strategies, we now construct infinite-memory strategies for arbitrary ECs.

Lemma 5. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP, let s_0 be a state in an EC U of MDP, and let $\vec{\nu} \geq \vec{0}$ be an expectation vector $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G} \downarrow U}^+(s_0)$ which is achievable while remaining in U. There exists a strategy $f \in \Delta(\mathcal{G})$ (not necessarily remaining in U) s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, f)$.*

Remark 5. The statement of the lemma holds even with f a *pure* strategy, by applying Remark 1 when constructing the expectation strategy f^{exp} below. However, randomized strategies suffice for our purposes.

The rest of this section is devoted to the proof of Lemma 5. We proceed by combining in a non-trivial way a strategy for the expectation with a strategy for the worst-case. Let f^{wc} be a worst-case strategy s.t. $\vec{0} \in \mathsf{WCSol}_{\mathcal{G}}^+(s, f^{wc})$ for every state s, which exists since the $\mathcal{G}$ is pruned. Let f^{exp} be a expectation strategy s.t. $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G} \downarrow U}^+(s_0, f^{exp})$. By Lemma 1, we can assume that f^{exp} is finite-memory and unichain. For technical reasons, it is convenient to assume that f^{exp} is finite-memory, even though we are going to construct a infinite-memory strategy. Moreover, since we are in an EC, we can further assume that f^{exp} achieves expectation $> \vec{\nu}$ from *every* state of the EC.

The idea is to play according to two different modes. In the first mode, we play according to f^{exp}, and in the second mode according to f^{wc}. We start in the first mode, and possibly go to the second mode according to certain conditions. This happens with a certain probability, which we call *switching probability*. Once in the second mode, we remain in the second mode. In order to achieve an expectation arbitrarily close to that achieved by f^{exp}, we need to be able to make the switching

probability arbitrarily small. At the same time, in order to ensure that the worst-case objective is satisfied, we need to guarantee that, when no switch occurs, the mean payoff is surely $> \vec{0}$. (If a switch occurs, the worst-case is satisfied by the definition of f^{wc}.) These two constraints are conflicting and make the construction of a combined strategy non-trivial.

The *combined* strategy f_K is parameterized by a natural number $K > 0$. In order to decide whether to switch to the second mode or not, we keep track of the total payoff since the beginning of the play as a vector in $\mathbb{Z}^d$. This value is unbounded in general, and this is explains why the strategy uses infinite memory. Let $\vec{N}_i$ be

$$\vec{N}_i = \frac{\vec{\nu} \cdot i \cdot K}{2}.$$

Thus, during the first mode the expected total payoff at the end of phase i is $> 2 \cdot \vec{N}_{i+1}$. The first mode is split into *phases*, each of length K. During phase $i \geq 0$, we play according to f^{exp} for at most K steps. There are two conditions that can trigger a switch to the second mode:

[Switching condition 1 (SC1)] If we are in phase $i \geq 1$ and the total payoff since the beginning of the play is not always $> \vec{N}_i$ during the current phase, then switch to f^{wc} permanently.

[Switching condition 2 (SC2)] If the total payoff since the beginning of the play is not $> 2 \cdot \vec{N}_{i+1}$ at the end of the current phase, then switch to f^{wc} permanently.

What it remains to do is to show that we can choose $K > 0$ in order to satisfy the BWC objective. First, we show that, for every choice of the parameter K, the combined strategy f_K guarantees the worst-case objective.

Proposition 5. *For every $K \in \mathbb{N}$ and state s_0 in the EC U, $\vec{0} \in \mathsf{WCSol}_{\mathcal{G}}^+(s_0, f_K)$.*

Proof. There are two cases to consider. If we ever switch to the second mode, then the run is eventually consistent with the worst-case strategy f^{wc}, which guarantees worst-case mean payoff $> \vec{0}$ (by prefix independence). Otherwise, assume that we never leave the first mode. During phase $i \geq 1$ the total payoff is always $> \vec{N}_i = \frac{\vec{\nu} \cdot i \cdot K}{2}$, and the total length of the play is at most $i \cdot K$. The average mean payoff during phase i is uniformly $> \frac{\vec{\nu}}{2}$. The limit inferior of the average mean payoff is also $> \frac{\vec{\nu}}{2} \geq \vec{0}$. $\square$

We conclude by showing that K can be chosen s.t. the combined strategy f_K achieves expected mean payoff $> \vec{\nu}$.

Lemma 6. *There exists $K \in \mathbb{N}$ s.t. $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}^+(s_0, f_K)$.*

Proof. We show that we can choose a $K > 0$ large enough s.t. the switching probability is negligible, and thus the impact of switching to the worst-case strategy f^{wc} on the expected mean payoff is also negligible. For now fix an arbitrary $K > 0$, and consider the Markov chain $\mathcal{G}[f_K]$. Let p_K be the probability to switch to the second mode due to SC1 in any phase $i \geq 1$, and let q_K be the probability to switch to the second mode

due to SC2 in any phase $i \geq 0$. Thus, with probability at most $1 - (1-p_K) \cdot (1-q_K)$ we switch to the second mode. By prefix independence of the mean payoff objective, the expected mean payoff achieved by f_K satisfies:

$$\mathbb{E}_{s_0, f_K}^{\mathcal{G}}[\mathsf{MP}] \geq \quad (1 - (1-p_K) \cdot (1-q_K)) \cdot \mathbb{E}_{s_0, f^{wc}}^{\mathcal{G}}[\mathsf{MP}] +$$
$$+ (1-p_K) \cdot (1-q_K) \cdot \mathbb{E}_{s_0, f^{exp}}^{\mathcal{G}}[\mathsf{MP}]$$

Since $\mathbb{E}_{s, f^{exp}}^{\mathcal{G}|U}[\mathsf{MP}] > \vec{\nu}$ by definition, it suffices to show that both probabilities p_K and q_K can be made arbitrarily small. We argue about them separately.

Let $p_{i,K}$ be the probability of switching to the second mode due to SC1 during phase $i \geq 1$, i.e, the probability that the total payoff goes below $\vec{N}_i$ in any component:

$$p_{i,K} = \mathbb{P}_{s_0}^{\mathcal{G}[f_K]}\left[\exists (K \cdot i \leq h < K \cdot (i+1)) \cdot \mathsf{TP}_h \not> \vec{N}_i\right]$$

Then, $p_K = p_{1,K} + (1-p_{1,K}) \cdot p_{2,K} + (1-p_{1,K}) \cdot (1-p_{2,K}) \cdot p_{3,K} + \ldots$, and thus $p_K \leq p_{1,K} + p_{2,K} + \ldots$. We claim the following exponential upper bound on $p_{i,K}$.

Claim 1. *There are rational constants a and b with $b < 1$ s.t., for every $i \geq 1$ and for sufficiently large K, $p_{i,K} \leq a \cdot b^{K \cdot i}$. Note that a and b do not depend neither on K, nor on i.*

By the claim, $p_K \leq a \cdot (b^K + b^{K \cdot 2} + \ldots) \leq a \cdot b^K / (1 - b^K)$, and thus $\lim_K p_K = 0$ since $b^K < 1$.

Let $q_{i,K}$ be the probability of switching to the second mode due to SC2 at the end of phase $i \geq 0$. Thus, $q_{i,K}$ is the probability that, at the end of phase i, the total payoff is less than $2 \cdot \vec{N}_{i+1} = \vec{\nu} \cdot (i+1) \cdot K$ in any component:

$$q_{i,K} = \mathbb{P}_{s_0}^{\mathcal{G}[f_K]}\left[\mathsf{TP}_{K \cdot (i+1)} \not> 2 \cdot \vec{N}_{i+1}\right]$$

We have $q_K = q_{0,K} + (1-q_{0,K}) \cdot q_{1,K} + (1-q_{0,K}) \cdot (1-q_{1,K}) \cdot q_{2,K} + \ldots$. We show $\lim_K q_K = 0$ as in the last paragraph, by the following claim.

Claim 2. *There exist rational constants a and b with $b < 1$ s.t., for every $i \geq 0$ and sufficiently large K, $q_{i,K} \leq a \cdot b^{K \cdot (i+1)}$. Note that a and b do not depend neither on K, nor on i.*

Both claims are proved in the appendix. $\square$

2) The general case: As in the synthesis for finite-memory strategies (cf. Sec. III-A), we reduce the infinite-memory BWC problem to the solution of a system of linear inequalities. The new system of equations T' is shown in Fig. 5. It is obtained as a modification of system T from the finite-memory case shown in Fig. 4: Specifically, T' is the same as T, except that Equations (A2), (B), and (C3) are interpreted w.r.t. MEC (instead of MWEC). The correctness of the reduction is stated in the lemma below.

Lemma 7. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP, let $\vec{\nu} \geq \vec{0}$, and let $s_0 \in S$. There exists a (possibly infinite-memory) strategy h s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h)$, if, and only if, the system T' has a non-negative solution.*

Proof sketch. The proof is analogous to the proof of Lemma 4. The crucial difference is that, by the modifications performed

$$1_{s_0}(s) + \sum_{(r,s)\in E} y_{rs} = \sum_{(s,t)\in E} y_{st} + y_s \qquad \forall s \in S \qquad\qquad\text{(A1)}$$

$$y_{st} = R(s)(t)\cdot\Big(\sum_{(r,s)\in E} y_{rs} - y_s\Big) \qquad \forall(s,t)\in E \text{ with } s\in S^R \qquad\qquad\text{(A1')}$$

$$\sum_{\text{MEC } U}\sum_{s\in U} y_s = 1 \qquad\qquad\qquad\qquad\text{(A2-bis)}$$

$$\sum_{s\in U} y_s = \sum_{(r,s)\in E\cap U\times U} x_{rs} \qquad \forall \text{ MEC } U \qquad\qquad\text{(B-bis)}$$

$$\sum_{(r,s)\in E} x_{rs} = \sum_{(s,t)\in E} x_{st} \qquad \forall s\in S \qquad\qquad\text{(C1)}$$

$$x_{st} = R(s)(t)\cdot\sum_{(r,s)\in E} x_{rs} \qquad \forall(s,t)\in E \text{ with } s\in S^R \qquad\qquad\text{(C1')}$$

$$\sum_{(s,t)\in E} x_{st}\cdot w(s,t)[i] > \vec{\nu}[i] \qquad \forall(1\le i\le n) \qquad\qquad\text{(C2)}$$

$$\sum_{(s,t)\in E\cap U\times U} x_{st}\cdot w(s,t)[i] > 0 \qquad \forall \text{ MEC } U, 1\le i\le n \qquad\qquad\text{(C3-bis)}$$

Fig. 5: Linear system T' for the BWC infinite-memory threshold problem.

to obtain T' from T, we obtain strategies which almost surely stay forever inside ECs, instead of WECs. Since we are allowed infinite-memory, we can approximate the BWC objective inside ECs by replacing Lemma 2 with Lemma 5. □

Running example. An infinite-memory strategy can benefit both from the expectation $(5, 15)$ in U and from $(15, 5)$ in V (which is not a WEC). By going to either EC with equal probability and playing according to a local BWC strategy, an infinite-memory strategy can ensure, for every $\varepsilon > 0$, the BWC threshold $((0,0);(10-\varepsilon, 10-\varepsilon))$.

3) Complexity: We obtain the following complexity result for the threshold problem for arbitrary controllers.

Theorem 6. *The multidimensional mean-payoff BWC threshold problem is* coNP-*complete.*

Proof. Pruning the game to remove states which are losing for the worst-case objective requires solving a multidimensional mean-payoff game, which is coNP-complete by Theorem 2. Then, by Lemma 7, it suffices to solve system L'. Notice that system L' is of polynomial size since there are only polynomially many *maximal* ECs. □

Again, it is the worst-case problem that dominates the complexity of the BWC problem. By restricting to essentially worst-case unidimensional thresholds we obtain a better complexity, as in Sec. III-A3.

Corollary 2. *The multidimensional mean-payoff BWC threshold problem w.r.t. essentially worst-case unidimensional thresholds is in* NP∩coNP.

This solves with optimal complexity the infinite-memory unidimensional BWC problem, which was left open in [3].

IV. BEYOND ALMOST-SURE SYNTHESIS

We introduce a natural relaxation of the BWC problem which enjoys a better complexity. Intuitively, we replace the worst-case objective in the BWC problem with a weaker almost sure objective. While the BWC problem is coNP-complete, we show that this relaxation can be solved in P, even in the multidimensional setting. A similar result has recently been obtained in [5]. Given an MDP $\mathcal{G}$, a starting state s_0 therein, and a Controller's strategy $f \in \Delta(\mathcal{G})$, the set of *almost sure achievable solutions for f*, denoted $\mathsf{ASSol}^+{}_{\mathcal{G}}(s_0, f)$, is the set of vectors $\vec{\mu} \in \mathbb{R}^d$ s.t. Controller can almost surely guarantee mean payoff $> \vec{\mu}$ when playing according to f, i.e., $\mathsf{ASSol}^+{}_{\mathcal{G}}(s, f) = \left\{ \vec{\mu} \in \mathbb{R}^d \;\middle|\; \mathbb{P}^{\mathcal{G}}_{s,f}[\mathsf{MP} > \vec{\mu}] = 1 \right\}$. The set of *beyond almost-sure achievable solutions for f*, denoted $\mathsf{BASSol}^+_{\mathcal{G}}(s_0, f)$, is the set of pairs of vectors $(\vec{\mu};\vec{\nu}) \in \mathbb{R}^{2d}$ s.t. Controller can almost surely guarantee mean payoff $> \vec{\mu}$ and achieve expected mean payoff $> \vec{\nu}$ when starting from s_0 and playing according to f, i.e.,

$$\mathsf{BASSol}^+_{\mathcal{G}}(s_0, f) = \left\{ (\vec{\mu};\vec{\nu}) \in \mathbb{R}^{2d} \;\middle|\; \begin{array}{c} \vec{\mu} \in \mathsf{ASSol}^+{}_{\mathcal{G}}(s_0, f) \\ \text{and} \\ \vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G}}(s_0, f) \end{array} \right\}$$

The set of *beyond almost-sure achievable solutions* is $\mathsf{BASSol}^+_{\mathcal{G}}(s_0) = \bigcup_{f\in\Delta(\mathcal{G})} \mathsf{BASSol}^+_{\mathcal{G}}(s_0, f)$. Given $(\vec{\mu};\vec{\nu}) \in \mathbb{R}^{2d}$ and a state s_0, the *beyond almost-sure threshold problem* asks whether $(\vec{\mu};\vec{\nu}) \in \mathsf{BASSol}^+_{\mathcal{G}}(s_0)$.

Remark 6. We assume w.l.o.g. that $\vec{\mu} = \vec{0}$ and $\vec{\nu} \ge \vec{0}$. The first condition is ensured by subtracting $\vec{\mu}$ everywhere. The second condition follows from the observation that, if the mean payoff is $> \vec{0}$ almost surely, then also the expectation is $> \vec{0}$ surely.

We observe that, inside an EC, there is no trade-off between the almost sure and the expectation objective.

Lemma 8. *Let $\mathcal{G}$ be a multidimensional mean-payoff MDP, let s_0 be a state in an EC U thereof, and let $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G} \downarrow U}(s_0)$ be an expectation achievable while remaining inside U. There exists a finite-memory strategy $g \in \Delta_F(\mathcal{G})$ s.t. $(\vec{\nu}; \vec{\nu}) \in \mathsf{BASSol}^+_{\mathcal{G} \downarrow U}(s_0, g)$ which also remains inside U.*

Proof. By Lemma 1, there exists a finite-memory strategy g s.t. $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G} \downarrow U}(s_0, g)$ and $\mathcal{G}[g]$ is unichain. Consequently, the mean payoff is $> \vec{\nu}$ almost surely. ☐

Thus, most of the effort goes in analyzing the general case. As in the BWC problem, we reduce the BAS problem to the solution of a system of linear inequalities. We assume that from state s_0 all ECs are reachable with positive probability. It turns out that the same system of equations T' used in the infinite-memory BWC threshold problem also solves the BAS problem. We obtain a better complexity since we do not require the MDP to be pruned (which avoids solving an expensive mean-payoff game).

Theorem 7. *The multidimensional mean-payoff BAS threshold problem is in* P.

Proof. The proof of correctness is the same as in Lemma 7, where Lemma 8 replaces Lemma 5 in the analysis of ECs. Crucially for complexity, we do not need to assume that the MDP is pruned. Therefore, system T' can be built (and solved) in P. Since Lemma 8 even yields finite-memory strategies inside an EC, the construction of Lemma 7 shows that finite-memory strategies suffice for the BAS threshold problem. (This relies on the strict BAS semantics. If non-strict inequalities are used, then the problem can still be solved in P but the construction above yields an infinite-memory strategy, and infinite-memory strategies are more powerful than finite-memory ones for the non-strict BAS problem; cf. also [5].) ☐

Running example. The BAS problem is strictly weaker than the BWC problem. Consider the MDP from Fig. 3 without the edge (u, t). This modification makes both states u and v losing for the worst-case, thus they are pruned away when solving the BWC problem (even with infinite memory). On the other hand, the mean payoff is almost surely $(5, 15)$ from V, and thus it satisfies the almost sure objective $> (0, 0)$. Therefore, for every $\varepsilon > 0$, we can achieve the BAS threshold $((0, 0); (10 - \varepsilon, 10 - \varepsilon))$ by going to t or u with equal probability.

V. CONCLUSIONS

In this paper, we studied the multidimensional generalization of the beyond worst-case problem introduced by Bruyère *et al.* [3]. We have provided tight coNP-completeness results for the this problem under both finite-memory and general strategies. Since multidimensional mean-payoff games are already coNP-complete, our upper bound shows that we can add a multidimensional expectation optimization objective on top of a worst-case requirement without a corresponding increase in complexity. Notice that, while infinite-memory strategies were known to be more powerful than finite-memory ones already in the unidimensional setting [3], the corresponding

synthesis problem was left open. Our results thus complete the complexity picture for this problem. Moreover, we showed that, when the worst-case objective is unidimensional, the complexity reduces to NP∩coNP, and this holds even for multidimensional expectations. This generalizes with optimal complexity the NP∩coNP upper bound for the unidimensional beyond worst-case problem [3]. From a practical point of view, our reductions to linear programming can be performed in pseudo-polynomial time by using the results of [15] for unidimensional mean-payoff games, and [16] for *fixed* number of dimensions. Furthermore, we introduced the beyond almost-sure problem as a natural relaxation of the beyond worst-case problem, by weakening the worst-case requirement to an almost-sure one. This natural relaxation enjoys a polynomial time solution and finite memory strategies always suffice. Moreover, our reduction to linear programming shows that the beyond almost-sure problem is amenable to be solved efficiently in practice, and thus it has the strongest appeal for practical applications.

REFERENCES

[1] T. M. Liggett and S. A. Lippman, "Stochastic games with perfect information and time average payoff," *SIAM Review*, vol. 11, no. 4, pp. 604–607, 1969. [Online]. Available: http://www.jstor.org/stable/2029090

[2] A. Ehrenfeucht and J. Mycielski, "Positional strategies for mean payoff games," *International Journal of Game Theory*, vol. 8, no. 2, pp. 109–113, 1979. [Online]. Available: http://dx.doi.org/10.1007/BF01768705

[3] V. Bruyère, E. Filiot, M. Randour, and J.-F. Raskin, "Meet Your Expectations With Guarantees: Beyond Worst-Case Synthesis in Quantitative Games," in *Proc. of STACS'14*, ser. LIPIcs, vol. 25, 2014, pp. 199–213.

[4] ——, "Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games," Tech. Rep., 09 2013. [Online]. Available: http://arxiv.org/abs/1309.5439

[5] K. Chatterjee, Z. Komárková, and J. Křetínský, "Unifying two views on multiple mean-payoff objectives in markov decision processes," *To appear in Proc. of LICS'15*.

[6] M. Puterman, *Markov Decision Processes - Discrete Stochastic Dynamic Programming*. John Wiley&Sons, Inc., 1994.

[7] Y. Velner and A. Rabinovich, "Church synthesis problem for noisy input," in *Proc. of FOSSACS'11*. Springer, 2011, pp. 275–289. [Online]. Available: http://dl.acm.org/citation.cfm?id=1987171.1987194

[8] Y. Velner, K. Chatterjee, L. Doyen, T. A. Henzinger, A. Rabinovich, and J.-F. Raskin, "The complexity of multi-mean-payoff and multi-energy games," Tech. Rep., 09 2012. [Online]. Available: http://arxiv.org/abs/1209.3234

[9] T. Brázdil, V. Brozek, K. Chatterjee, V. Forejt, and A. Kucera, "Two views on multiple mean-payoff objectives in markov decision processes," *LMCS*, vol. 10, no. 1, 2014. [Online]. Available: http://dx.doi.org/10.2168/LMCS-10(1:13)2014

[10] M. Randour, J.-F. Raskin, and O. Sankur, "Percentile queries in multi-dimensional markov decision processes," in *Proc. of CAV'15*, to appear.

[11] U. Zwick and M. Paterson, "The complexity of mean payoff games on graphs," *TCS*, vol. 158, no. 1–2, pp. 343–359, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0304397595001883

[12] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, 1995.

[13] L. De Alfaro, "Formal verification of probabilistic systems," Ph.D. dissertation, Stanford University, Stanford, CA, USA, 1998.

[14] M. Jurdziński, "Deciding the winner in parity games is in UP∩co-UP," *Information Processing Letters*, vol. 68, no. 3, pp. 119–124, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020019098001501

[15] L. Brim, J. Chaloupka, L. Doyen, R. Gentilini, and J. Raskin, "Faster algorithms for mean-payoff games," *Formal Methods in System Design*, vol. 38, no. 2, pp. 97–118, 2011.

[16] K. Chatterjee and Y. Velner, "Hyperplane separation technique for multidimensional mean-payoff games," in *Proc. of CONCUR'13*, ser. LNCS, vol. 8052. Springer, pp. 500–515.

A. Hoeffding-style bounds

In this section, we prove an Hoeffding-style bound for multidimensional Markov chains which will be used repeatedly in later proofs. Let $\mathcal{G}$ be a Markov chain which is unichain. Recall that a Markov chain is unichain when it consists of transient states and a unique bottom strongly connected component. Thus, any run in $\mathcal{G}$ will be trapped almost surely in the bottom component, and the mean payoff will be almost surely equal to the expected mean payoff. Moreover, the expected mean payoff is the same from every starting state of $\mathcal{G}$. Below, we present a bound on the probability that the mean payoff deviates from the expected mean payoff for sufficiently long runs. Let $d \geq 1$ be the dimension of the weights in the Markov chain $\mathcal{G}$, and let $\vec{\nu}$ be the expected mean-payoff vector from any state in $\mathcal{G}$.

Lemma 9. *For any $\delta > 0$, there exists $K_0 = O(\frac{1}{\delta^2}) \in \mathbb{N}$ and constants $a, b > 0$ s.t., for all $K \geq K_0$ and state s,*

$$\mathbb{P}_s^{\mathcal{G}} \left[\exists (1 \leq j \leq d) \cdot |(\mathsf{MP}_K - \vec{\nu})[j]| \geq \delta \right] \leq \mathcal{G}(K, \delta) := 2^d \cdot a \cdot e^{-b \cdot K \cdot \delta^2} .$$

Moreover, a and b are polynomial in the parameters of the Markov chain, a is exponential in δ, and K_0 is polynomial in the size of the Markov chain and in the largest absolute weight W (and thus exponential in its encoding).

We prove the lemma above by reducing to the unidimensional case $d = 1$. The latter case was already dealt with in [4, Lemma 9], which in turns relies on [17, Proposition 2].

Lemma 10 (cf. [4, Lemma 9]). *For any $\delta > 0$, there exist $K_0 = O(\frac{1}{\delta}) \in \mathbb{N}$ and constants $a, b > 0$ s.t., for all $K \geq K_0$ and state s,*

$$\mathbb{P}_s^{\mathcal{G}} \left[|\mathsf{MP}_K - \nu| \geq \delta \right] \leq \mathcal{F}_{a,b}(K, \delta) := a \cdot e^{-b \cdot K \cdot \delta^2} .$$

Moreover, a and b are polynomial in the parameters of the Markov chain, a is exponential in δ. and, K_0 is polynomial in the size of the Markov chain and in the largest weight W (and thus exponential in its encoding).

Proof of Lemma 9. In the following, fix an error $\delta > 0$ and a number of steps $K > 0$. For a component $1 \leq j \leq d$, we say that a run π is *j-bad* iff the j-th component of the mean payoff deviates from $\vec{\nu}[j]$ at least by δ after K steps, i.e., if

$$|(\mathsf{MP}(\pi(K)) - \vec{\nu})[j]| \geq \delta$$

and π is *j-good* otherwise. Moreover, we say that π is *bad* if it is j-bad for some $1 \leq j \leq d$, and we say that π is *good* it is j-good for every $1 \leq j \leq d$. In other words, π is good if for every component j, $|(\mathsf{MP}(\pi(K)) - \vec{\nu})[j]| < \delta$.

For a fixed dimension $1 \leq j \leq d$, we are in the unidimensional case, dealt with in Lemma 10. Let j be a fixed dimension. By Lemma 10, there exist constants $a_j, b_j > 0$, and $K_j = O(\frac{1}{\delta}) \in \mathbb{N}$ s.t., for every $K \geq K_j$ and state s, $\mathcal{F}_{a_j,b_j}(K, \delta)$ is an upper bound on the probability that π is j-bad. We want to choose uniform $a, b > 0$ and K_0 s.t.

$$\mathcal{F}_{a_j,b_j}(K, \delta) \leq \mathcal{F}_{a,b}(K, \delta) < 1$$

for every $K \geq K_0$. To this end, let

$$a := \max_j a_j,$$
$$b := \min_j b_j, \text{ and}$$
$$K_0 := \max\{\max_j K_j, \left\lceil \frac{\ln a}{b \cdot \delta^2} \right\rceil + 1\} .$$

Note that $K_0 = O(\frac{1}{\delta^2})$. Then, $1 - \mathcal{F}_{a,b}(K, \delta)$ is a lower bound on the probability that π is j-good, for any fixed j and $K \geq K_0$. Then, $(1 - \mathcal{F}_{a,b}(K, \delta))^d$ is a lower bound on the probability that π is good. We derive the following simple lower bound on

the latter quantity:

$$(1 - \mathcal{F}_{a,b}(K,\delta))^d = \sum_{i=0}^{d} \binom{d}{i}(-\mathcal{F}_{a,b}(K,\delta))^i =$$

$$= 1 + \sum_{i=1}^{d} \binom{d}{i}(-\mathcal{F}_{a,b}(K,\delta))^i$$

$$\geq 1 - \sum_{i=1}^{d} \binom{d}{i}(\mathcal{F}_{a,b}(K,\delta))^i$$

$$\geq 1 - \sum_{i=1}^{d} \binom{d}{i}\mathcal{F}_{a,b}(K,\delta)$$

$$\geq 1 - 2^d \cdot \mathcal{F}_{a,b}(K,\delta)$$

Finally, $1 - (1 - \mathcal{F}_{a,b}(K,\delta))^d$ is an upper bound on the probability that π is bad. We define

$$\mathcal{G}(K,\delta) := 2^d \cdot \mathcal{F}_{a,b}(K,\delta) = 2^d \cdot a \cdot e^{-b \cdot K \cdot \delta^2}$$

By the inequality above, $\mathcal{G}(K,\delta) \geq 1 - (1 - \mathcal{F}_{a,b}(K,\delta))^d$, and thus $\mathcal{G}(K,\delta)$ is an upper bound on the probability that a period is bad for every $K \geq K_0$. □

B. Mean-payoff value function

We make a couple of simple observations on the relationship between the mean-payoff value function and the total-payoff value function.

Lemma 11. *Let π be a play in a graph G.*
1) If $\mathsf{MP}(\pi) > 0$, then $\mathsf{TP}(\pi) = +\infty$. Therefore, if $\mathsf{TP}(\pi) < +\infty$ (with possibly $\mathsf{TP}(\pi) = -\infty$), then $\mathsf{MP}(\pi) \leq 0$.
2) Similarly, if $\mathsf{MP}(\pi) < 0$, then $\mathsf{TP}(\pi) = -\infty$.
3) There exists a play π_0 in a finite graph G_0 s.t. $\mathsf{MP}(\pi_0) = 0$, but $\mathsf{TP}(\pi_0) = +\infty$.
4) There exists a play π_1 in a finite graph G_1 s.t. $\mathsf{MP}(\pi_1) = 0$, but $\mathsf{TP}(\pi_1) = -\infty$.

Proof. We first prove Point 1). Assume $\mathsf{MP}(\pi) = a$ for some $a > 0$. By the definition of $\liminf$, for every $\varepsilon > 0$ there exists m_0 s.t. for every $m \geq m_0$ we have $\mathsf{MP}(\pi(m)) \geq a - \varepsilon$. To show $\mathsf{TP}(\pi) = +\infty$ we show that for every bound $b > 0$, there exists n_0 s.t. for every $n \geq n_0$, $\mathsf{TP}(\pi(n)) \geq b$. Let $b > 0$. If we take $\varepsilon := a/2$ in the definition above, we have that there exists m_0 s.t. for every $m \geq m_0$ $\mathsf{MP}(\pi(m)) \geq a/2 > 0$. We take $n_0 := \max\{m_0, 2b/a\}$, and let $n \geq n_0$. Since $\mathsf{TP}(\pi(n)) = n \cdot \mathsf{MP}(\pi(n))$, we have $\mathsf{TP}(\pi(n)) \geq na/2 \geq b$, where the latter inequality follows from the definition of n_0.

The proof of Point 2) is analogous to the proof of Point 1).

For Point 3), consider a play π_0 inducing the following sequence of payoffs:

$$\underbrace{1}_{2^0}\underbrace{10}_{2^1}\underbrace{1000}_{2^2}\underbrace{10000000}_{2^3}\cdots$$

i.e., the n-th payoff is 1 if n is a power of 2, and 0 otherwise. Then, $\mathsf{TP}(\pi_0(n)) = k$ where k is the largest exponent s.t. $2^k \leq n$, i.e., $k = \lfloor \lg n \rfloor$. Thus, $\mathsf{TP}(\pi_0) = +\infty$. However, $\mathsf{MP}(\pi_0(n)) = \frac{\lfloor \lg n \rfloor}{n}$ goes to 0 as n goes to $+\infty$. Thus, $\mathsf{MP}(\pi_0) = 0$. Point 4) is proved analogously by taking the sequence

$$\underbrace{(-1)}_{2^0}\underbrace{(-1)0}_{2^1}\underbrace{(-1)000}_{2^2}\underbrace{(-1)0000000}_{2^3}\cdots$$

□

As an application of Lemma 10, we show that if the total payoff is $-\infty$ almost surely, then the mean payoff is strictly negative almost surely. This contrasts with Point 4) in Lemma 11, which showed that there are infinite runs with total payoff equal to $-\infty$, but which have nonetheless zero mean payoff. (Notice that the infinite play constructed in the proof of the latter lemma with this property was non-periodic.) We use the lemma below later in the proof of Lemma 21.

Lemma 12. *Let $\mathcal{G}$ be a Markov chain. For every state s_0,*

$$\mathbb{P}_{s_0}^{\mathcal{G}}[\mathsf{TP} = -\infty] = \mathbb{P}_{s_0}^{\mathcal{G}}[\mathsf{MP} < 0] .$$

In particular, if the mean payoff is non-negative almost surely, then the total payoff is $> -\infty$ almost surely.

$$\sum_{s \in S} x_s = 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(EC-1)}$$

$$x_s = \sum_{(r,s) \in E} x_{rs} \qquad\qquad\qquad \forall s \in S \qquad\qquad\qquad\qquad \text{(EC-IN)}$$

$$x_s = \sum_{(s,t) \in E} x_{st} \qquad\qquad\qquad \forall s \in S \qquad\qquad\qquad\qquad \text{(EC-OUT)}$$

$$x_{st} = R(s)(t) \cdot x_s \qquad\qquad\qquad \forall (s,t) \in E \text{ with } s \in S^R \qquad \text{(EC-RAND)}$$

$$\sum_{(s,t) \in E} x_{st} \cdot w(s,t)[i] \geq \vec{\nu}[i] \qquad\qquad \forall (1 \leq i \leq d) \qquad\qquad\qquad \text{(EC-MP)}$$

Fig. 6: System of linear inequalities for the expectation problem inside an EC.

Proof. Fix a state s_0. Let $p = \mathbb{P}_{s_0}^{\mathcal{G}}[\mathsf{MP} < 0]$, $q = \mathbb{P}_{s_0}^{\mathcal{G}}[\mathsf{TP} = -\infty]$, and, for a BSCC B, let p_B be the probability of reaching B. In a BSCC B, MP and TP take value equal to their respective expectations almost surely, and this value is the same from every state in the BSCC. Let this value be $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}]$ and $\mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}]$, respectively. We thus have the following decomposition:

$$p = \sum_{\text{BSCC } B \text{ s.t. } \mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}]<0} p_B, \text{ and}$$

$$q = \sum_{\text{BSCC } B \text{ s.t. } \mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}]=-\infty} p_B \ .$$

It suffices to show that $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] < 0$ if, and only if, $\mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}] = -\infty$ for all BSCC's B.

If $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] < 0$, then $\mathsf{MP} < 0$ almost surely since B is a BSCC. By Point 2) of Lemma 11, $\mathsf{MP} < 0$ implies $\mathsf{TP} = -\infty$ surely, and thus $\mathsf{TP} = -\infty$ holds almost surely, and consequently $\mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}] = -\infty$.

For the other direction, let $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] \geq 0$. If $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] > 0$, then by Point 1) of Lemma 11 and reasoning as above, we obtain $\mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}] = +\infty$. It remains to prove the case $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] = 0$. This makes use of the bound provided by Lemma 10, and it does not hold in a non-probabilistic setting (cf. the counter-example in Point 4) of Lemma 11). Assume $\mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] = 0$. We prove $\mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}] > -\infty$. For every $s_1 \in B$ and K, we have

$$\mathbb{P}_{s_1}^{\mathcal{G}}[\mathsf{TP}_K \leq -K] \leq \mathbb{P}_{s_1}^{\mathcal{G}}[\mathsf{MP}_K \leq -1]$$
$$\leq \mathbb{P}_{s_1}^{\mathcal{G}}[|\mathsf{MP}_K| \geq 1]$$
$$\leq \mathcal{F}_{a,b}(K,1)$$

where the last inequality follows from Lemma 10 applied with $\nu = \mathbb{E}_B^{\mathcal{G}}[\mathsf{MP}] = 0$, $\delta = 1$, for some constants $a, b > 0$ and for K sufficiently large. Since $\mathcal{F}_{a,b}(K,1) \to 0$ as $K \to \infty$, we have that $\mathbb{P}_{s_1}^{\mathcal{G}}[\mathsf{TP} = -\infty] = 0$, which implies $\mathbb{E}_B^{\mathcal{G}}[\mathsf{TP}] > -\infty$. $\square$

C. Finite-memory synthesis in an EC

In this section, we show that achievable values can be approximated by randomized finite-memory strategies in ECs, with the further property that the induced (finite) Markov chain is unichain, i.e., it contains exactly one BSCC.

Lemma 1. *Let $\mathcal{G}$ be a multidimensional mean-payoff MDP, let s_0 be a state in an EC U thereof, and let $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}|U}^+(s_0)$ be an expectation vector achievable by remaining inside U. There exists a* finite-memory unichain *strategy $g \in \Delta_F(\mathcal{G})$ achieving the same expectation $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}|U}^+(s_0, g)$.*

We prove this result as follows. First, in Sec. A-C1 we characterize the set of of achievable vectors as non-negative solutions to a linear programming problem (in the spirit of [9]). This yields a natural decomposition of the EC into several SCCs. For each such SCC, we construct in Sec. A-C2 a *randomized memoryless* "local strategy" achieving a corresponding "local expectation". No approximation error is introduced in this step. Then, in Sec. A-C3 we combine those "local strategies" into a randomized finite-memory "global strategy" approximating the expectation. This second step uses the fact that in an EC all states are inter-reachable (under some strategy), and thus we can cycle through all the "local strategies" for the appropriate fraction of time. This step introduces an approximation error, due to the cost of moving from a SCC to the next one. However, by using larger amounts of finite memory, we can make this error arbitrarily small.

1) Decomposition in SCCs: In the following, let $\mathcal{G} = (G, S^0, S^R, R)$ with $G = (d, S, E, w)$ be a fixed MDP. W.l.o.g. we assume that $\mathcal{G}$ is reduced to a single EC S. Let $\vec{v} \in \mathbb{Q}^d$ be an expected-value achievable vector. Consider the linear program $A_{\vec{v}}$ of Fig. 6. (Cf. [9] for a similar linear program in the more general case where the MDP is not just an EC.). We use the linear program $A_{\vec{v}}$ to obtain the long-run "frequencies" of edges guaranteeing mean payoff $\vec{v}$. For each state $s \in S$, we have a variable x_s representing the long-run probability to be in s, and, for each edge $(s, t) \in E$, we have a variable x_{st} for the long-run probability of taking edge (s, t). In the following, let

$$\mathsf{ExpSol}_{\mathcal{G}}(s_0, f) = \{\vec{v} \in \mathbb{R}^d \mid \mathbb{E}_{s_0, f}^{\mathcal{G}}[\mathsf{MP}] \geq \vec{v}\} \ ,$$

$$\mathsf{ExpSol}_{\mathcal{G}}(s_0) = \bigcup_{f \in \Delta(\mathcal{G})} \mathsf{ExpSol}_{\mathcal{G}}(s_0, f) \ .$$

The lemma below shows that $A_{\vec{v}}$ has a non-negative solution if $\vec{v}$ is achievable in $\mathcal{G}$. Correctness follows directly from the analysis of [9]. The complexity of the solution follows from [18, Theorem 10.1]. In the statement below, recall that W is the maximum absolute value of any weight in $\mathcal{G}$, and Q is be the largest denominator of any probability appearing therein.

Lemma 13. *Let $\mathcal{G}$ be a multidimensional mean payoff MDP reduced to a single EC S, let $s_0 \in S$ be a state therein, and let $\vec{v} \in \mathsf{ExpSol}_{\mathcal{G}}(s_0)$ be an achievable value for the expectation. Then, the system $A_{\vec{v}}$ has a non-negative solution of size[3] bounded by a polynomial in W and Q.*

Let $\{\tilde{x}_{st}\}_{(s,t) \in E}, \{\tilde{x}_s\}_{s \in S}$ be a non-negative solution to $A_{\vec{v}}$ of complexity polynomial in W and Q. This allows us to perform the following decomposition of $\mathcal{G}$ into strongly connected components. Let $S_{>0}$ be the set of states visited with (strictly) positive long-run average probability, and let $E_{>0}$ be the set of edges visited with (strictly) positive long-run average probability:

$$S_{>0} = \{s \in S \mid \tilde{x}_s > 0\}$$

$$E_{>0} = \{(s, t) \in E \mid \tilde{x}_{st} > 0\}$$

By the flow conditions of $A_{\vec{v}}$, $E_{>0} = E \cap (S_{>0} \times S_{>0})$, i.e., positive edges are exactly those connecting positive states. States in $S_{>0}$ can be partitioned into maximal strongly connected components $\{S_1, \ldots, S_k\}$ (w.r.t. $E_{>0}$), such that there is no positive edge between different components. Since states in S_i have at least one successor (as $S_i \subseteq S_{>0}$) and all successors are in fact inside S_i, we have that S_i is an EC in $\mathcal{G}$. Let $E_i = E_{>0} \cap S_i \times S_i$ be the restriction of $E_{>0}$ to S_i, let $G_i = (d, S_i, E_i, w)$ be the corresponding graph, and let $\mathcal{G}_i = (G_i, S_i^0, S_i^R, R)$ be the resulting MDP, where $S_i^0 = S^0 \cap S_i$ and $S_i^R = S^R \cap S_i$.

For each $i \in \{1, \ldots, k\}$, let $x_i > 0$ be the total long-run average probability of being in S_i, and let $\vec{v}_i$ be the expected mean payoff vector achieved when starting from (anywhere) in component S_i:

$$x_i = \sum_{s \in S_i} \tilde{x}_s > 0 \tag{3}$$

$$\vec{v}_i = \frac{1}{x_i} \cdot \sum_{(s,t) \in E_i} \tilde{x}_{st} \cdot w(s, t) \tag{4}$$

Since component S_i cannot be left, it is reached with probability x_i, and thus the expected mean payoff is $\sum_{i=1}^{k} x_i \cdot \vec{v}_i \geq \vec{v}$.

Remark 7. This analysis immediately yields a 2-memory *randomized* strategy achieving expected mean payoff $\vec{v}$. Such a strategy goes to SCC S_i with probability x_i, and then plays edge (s, t) with probability $\tilde{x}_{st}/\tilde{x}_s$. Two memory states are required to discriminate the two phases. However, such a strategy is not unichain in general, which is what we aim at in this section.

We design a randomized finite memory unichain strategy that plays edges $(s, t) \in E_{>0}$ with approximate long-run average frequency $\tilde{x}_{st}$, in order to have mean payoff close to $\vec{v}$. We do this in two steps. First, in Sec. A-C2 we design, for each SCC S_i, a randomized memoryless "local strategy" g_i which plays edge $(s, t) \in E_i$ with long-run average frequency $\tilde{x}_{st}/x_i$ when started inside S_i. Then, in Sec. A-C3 we combine those g_i's into a global strategy that spends in each S_i's an approximate long-run fraction of time x_i. By using larger amounts of memory, the error in this approximation can be made arbitrarily small.

2) Inside a SCC S_i (Strategy g_i): For each SCC S_i, let g_i be the randomized memoryless strategy that plays edge $(s, t) \in E_i$ with $s \in S_i^0$ with probability $\tilde{x}_{st}/\tilde{x}_s$. Thus, in $\mathcal{G}_i[g_i]$ edge $(s, t) \in E_i$ is visited for a long-run proportion of time $\tilde{x}_{st}/x_i$.

Lemma 14. *$\mathcal{G}_i[g_i]$ is recurrent, and for every state $s_0 \in S_i$,*

$$\mathbb{E}_{s_0, g_i}^{\mathcal{G}_i}[\mathsf{MP}] = \vec{v}_i$$

---

[3]The size of a non-negative rational number $x = p/q$ with $p, q \in \mathbb{N}$ relatively prime, and $q > 0$, is the size of their bit representation, $1 + \lceil \log_2(p + 1) \rceil + \lceil \log_2(q + 1) \rceil$; cf. [18, Section 3.2].
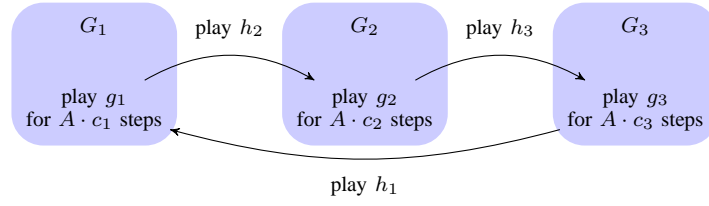
Fig. 7: The global strategy g_A

In the following lemma, we show the mean payoff obtained by playing according to g_i for sufficiently long time is close to $\vec{\nu}_i$ with high probability. The constant L_0 in the statement of the lemma does not depend on S_i, thus the guarantee holds in every component S_i. The lemma follows from a Hoeffding-style analysis.

Lemma 15. *For every $\delta > 0$, there exists $L_0 = O(\frac{1}{\delta^2}) \in \mathbb{N}$ and constants $a, b > 0$ s.t., for every component S_i, for every $L \geq L_0$ and for every state $s_0 \in S_i$,*

$$\mathbb{P}^{\mathcal{G}_i}_{s_0, g_i} \left[\exists (1 \leq j \leq d) \cdot |(\mathsf{MP}_L - \vec{\nu}_i)[j]| \geq \delta \right] \leq \mathcal{G}(L, \delta) := 2^d \cdot a \cdot e^{-b \cdot L \cdot \delta^2} \; .$$

Proof. The lemma follows from an application of Lemma 9 to each component S_i separately, and then by aggregating the constants. More precisely, for each irreducible (and thus unichain) Markov chain $\mathcal{G}_i[g_i]$, $1 \leq i \leq k$, Lemma 9 provides L_i (called K_0 in the lemma) and constants $a_i, b_i > 0$ s.t. for every $L \geq L_i$ and state $s_0 \in S_i$,

$$\mathbb{P}^{\mathcal{G}_i}_{s_0, g_i} \left[\exists (1 \leq j \leq d) \cdot |(\mathsf{MP}_L - \vec{\nu}_i)[j]| \geq \delta \right] \leq \mathcal{G}(L, \delta) := 2^d \cdot a_i \cdot e^{-b_i \cdot L \cdot \delta^2} \; .$$

Just take $L_0 := \max\{L_1, \ldots, L_k\}$, $a := \max\{a_1, \ldots, a_k\}$, and $b := \min\{b_1, \ldots, b_k\}$ to satisfy the claim. $\square$

3) Across SCCs (The global strategy): We now combine the local strategies g_i's in order to achieve approximate expected mean payoff $\vec{\mu}$ with finite memory. For each SCC S_i, let h_i be a memoryless strategy ensuring that S_i is reached almost surely from any state in $\mathcal{G}$. The strategy g_A is parametrized by a natural number $A > 0$. Assume that x_i is of the form $x_i = a_i/b_i$, with $a_i, b_i \in \mathbb{N}$ relatively prime, $b_i > 0$, let $b = \mathrm{lcm}\{b_1, \ldots, b_k\}$, and let $c_i = b \cdot x_i$. Note that c_i is a natural number. Intuitively, g_A works in k different stages. In stage $i \in \{1, \ldots, k\}$, g_A does the following:

(a) Play h_i to reach S_i almost surely.
(b) Once in S_i, switch to strategy g_i for $A \cdot c_i$ steps. Then, switch to stage $(i \mod k) + 1$ and go to *(a)*.

A full repetition of stages $\{1, \ldots, k\}$ is called a *phase*. Intuitively, g_A spends a proportion of time x_i in S_i in the limit, and, while the game stays in S_i, g_A plays according to g_i. Recall that g_i is memoryless.

Remark 8. Strategy g_A can be implemented with memory bounded by $k \cdot A \cdot b$. Notice that in both *(a)* and *(b)*, g_A plays according to a memoryless strategy, and no memory is needed to distinguish *(a)* from *(b)* since it suffices to look at the current state. Since the size of the binary representation of b is polynomial in W and Q (cf. Lemma 13), strategy g_A uses memory exponential in W and Q, and linear in n and A, where n is the number of states of $\mathcal{G}$.

We show that for any additive error $\varepsilon > 0$, we can play each stage sufficiently long (by increasing the parameter A) s.t. the probability of deviating from the expected mean payoff $\vec{\nu}$ by more than ε in any component is small.

Lemma 16. *For any achievable vector $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}(s_0)$ and $\varepsilon > 0$, there exists an $A_\varepsilon \in \mathbb{N}$ $(= O(\frac{1}{\varepsilon}))$ s.t., for every $A \geq A_\varepsilon$, $(\vec{\nu} - \vec{\varepsilon}) \in \mathsf{ExpSol}_{\mathcal{G}}(s_0, g_A)$.*

Proof. We begin by analysing the expected mean payoff of g_A over a single phase. Let $\vec{e}$ be the expected mean payoff of a single phase. We prove that for every $\varepsilon > 0$ there exists A large enough s.t. strategy g_A achieves at least expected mean payoff $\vec{\nu} - \vec{\varepsilon}$ over a single phase.

Let l be an upper bound on the expected length of periods of type *(a)*. That a finite such l exists can be seen as follows. Formally, let $H_{s,i}$ be the random variable that returns the *first hitting time* of the set S_i when starting from $s \in S$, that is the number of steps to reach S_i. Let $l_i = \max_{s \in S} \mathbb{E}^{\mathcal{G}}_{s, h_i}[H_{s,i}]$ be the worst expected first hitting time of S_i from any state in S when playing according to the memoryless strategy h_i (which reaches S_i almost surely), and take $l = \sum_{i=1}^{k} l_i$. By the definition of h_i and standard results about hitting times, the value l is finite.

The length of a period of type *(b)* at stage i is $A \cdot c_i$, thus the total length of periods of type *(b)* over a single phase is $\sum_{i=1}^{k} A \cdot c_i = A \cdot b$. Since l (computed above) is the expected length of periods of type *(a)* over a single phase, the expected length of a single phase is at most $A \cdot b + l$.

Let $\vec{e}_{(b),i}$ be the expected mean payoff of a period of type *(b)* at stage i. Apply Lemma 15 with $\delta := \varepsilon/8 > 0$, and let L_0 as given by the lemma. Thus, for every component S_i, if we play g_i for time $L \geq L_0$, we have the following lower bound on $\vec{e}_{(b),i}$:

$$\vec{e}_{(b),i} \geq (1 - \mathcal{G}\,(L, \varepsilon/8)) \cdot (\vec{\nu}_i - \vec{\varepsilon}/8) + \mathcal{G}\,(L, \varepsilon/8) \cdot (-\vec{W})$$

where W is the largest absolute value of weights in $\mathcal{G}$ and $\vec{W} = (W, \ldots, W) \in \mathbb{R}^d$. Moreover, since $\mathcal{G}\,(L, \varepsilon/8) \to 0$ as $L \to \infty$, there exists an $L_* \geq L_0$ s.t., for every $L \geq L_*$, $(1 - \mathcal{G}\,(L, \varepsilon/8)) \cdot (\vec{\nu}_i - \vec{\varepsilon}/8) + \mathcal{G}\,(L, \varepsilon/8) \cdot (-\vec{W}) \geq \vec{\nu}_i - \vec{\varepsilon}/4$, and thus

$$\vec{e}_{(b),i} \geq \vec{\nu}_i - \vec{\varepsilon}/4 \tag{5}$$

for every $L \geq L_*$. We derive a precise bound for L_*.

$$(1 - \mathcal{G}\,(L, \varepsilon/8)) \cdot (\vec{\nu}_i - \vec{\varepsilon}/8) + \mathcal{G}\,(L, \varepsilon/8) \cdot (-\vec{W}) \geq \vec{\nu}_i - \vec{\varepsilon}/4$$

$$\text{if} \quad \vec{\nu}_i - \vec{\varepsilon}/8 - \mathcal{G}\,(L, \varepsilon/8) \cdot (\vec{\nu}_i - \vec{\varepsilon}/8 + \vec{W}) \geq \vec{\nu}_i - \vec{\varepsilon}/4$$

$$\text{if} \quad \vec{\varepsilon}/8 - \mathcal{G}\,(L, \varepsilon/8) \cdot (\vec{\nu}_i - \vec{\varepsilon}/8 + \vec{W}) \geq \vec{0}$$

$$\text{if} \quad \varepsilon/8 - \mathcal{G}\,(L, \varepsilon/8) \cdot (\nu_i^{\max} - \varepsilon/8 + W) \geq 0$$

$$\text{if} \quad \mathcal{G}\,(L, \varepsilon/8) \leq \frac{\varepsilon}{8\nu_i^{\max} - \varepsilon + 8W}$$

$$\text{if} \quad a \cdot 2^d \cdot e^{-b \cdot L \cdot \varepsilon^2/64} \leq \frac{\varepsilon}{8\nu_i^{\max} - \varepsilon + 8W}$$

$$\text{if} \quad e^{-b \cdot L \cdot \varepsilon^2/64} \leq \frac{1}{a \cdot 2^d} \cdot \frac{\varepsilon}{8\nu_i^{\max} - \varepsilon + 8W}$$

$$\text{if} \quad -b \cdot L \cdot \varepsilon^2/64 \leq \ln\left(\frac{1}{a \cdot 2^d} \cdot \frac{\varepsilon}{8\nu_i^{\max} - \varepsilon + 8W}\right)$$

$$\text{if} \quad L \geq \frac{64}{b \cdot \varepsilon^2} \cdot \ln \frac{a \cdot 2^d \cdot (8\nu_i^{\max} - \varepsilon + 8W)}{\varepsilon} =$$

$$= \frac{64}{b \cdot \varepsilon^2}\,(\ln a + d \cdot \ln 2 + \ln(8\nu_i^{\max} - \varepsilon + 8W) - \ln \varepsilon)$$

where $\nu_i^{\max} = \max\{\vec{\nu}_i[1], \ldots, \vec{\nu}_i[d]\}$ is the largest component of $\vec{\nu}_i$ and $\mathcal{G}\,(L, \varepsilon/8) = a \cdot 2^d \cdot e^{-b \cdot L \cdot \varepsilon^2/64}$. Thus, take $L_* := \frac{64}{b \cdot \varepsilon^2}\,(\ln a + d \cdot \ln 2 + \ln(8\nu_i^{\max} - \varepsilon + 8W) - \ln \varepsilon)$. Notice that $L_* = O(\frac{1}{\varepsilon})$ (since a is exponential in ε).

Therefore, we stay in stage i at least L_* number of steps, which implies that we should have $A \cdot c_i \geq L_*$ for every $1 \leq i \leq k$.

Assumption 1. $A \geq A_0 := \max\{L_*, L_0\} = O(\frac{1}{\varepsilon})$.

By the definition of L_0, A_0 is exponential in n (the number of states of $\mathcal{G}$), and polynomial in W and Q.

For a period of type *(b)* in stage i, the expected total payoff is $A \cdot c_i \cdot \vec{e}_{(b),i}$. The expected total payoff of all periods of type *(a)* during a single phase is at least $-\vec{W} \cdot l$. Thus,

$$\vec{e} \geq \frac{A \cdot c_1 \cdot \vec{e}_{(b),1} + \cdots + A \cdot c_k \cdot \vec{e}_{(b),k} - \vec{W} \cdot l}{A \cdot b + l} =$$

$$= \frac{x_1 \cdot \vec{e}_{(b),1} + \cdots + x_k \cdot \vec{e}_{(b),k}}{1 + \frac{l}{A \cdot b}} - \frac{\vec{W} \cdot l}{A \cdot b + l}$$

First, we choose A large enough s.t. (A) $\frac{\vec{W} \cdot l}{A \cdot b + l} \leq \frac{\vec{\varepsilon}}{2}$:

$$\frac{\vec{W} \cdot l}{A \cdot b + l} \leq \frac{\vec{\varepsilon}}{2} \quad \text{iff} \quad \frac{W \cdot l}{A \cdot b + l} \leq \frac{\varepsilon}{2}$$

$$\text{iff} \quad 2 \cdot W \cdot l \leq \varepsilon \cdot (A \cdot b + l) = (\varepsilon \cdot b) \cdot A + \varepsilon \cdot l$$

$$\text{iff} \quad A \geq \frac{l \cdot (2 \cdot W - \varepsilon)}{\varepsilon \cdot b}$$

Assumption 2. $A \geq A_1 := \frac{l \cdot (2 \cdot W - \varepsilon)}{\varepsilon \cdot b} = O(\frac{1}{\varepsilon})$.

When $A \geq A_1$ we also have (B) $\vec{e}_{(b),i}/(1 + \frac{l}{A \cdot b}) \geq \vec{\nu}_i - \frac{\vec{\varepsilon}}{2}$, for every $1 \leq i \leq k$. Indeed,

$$\vec{e}_{(b),i}/\left(1 + \frac{l}{A \cdot b}\right) \geq \vec{\nu}_i - \frac{\vec{\varepsilon}}{2} \quad \text{if} \quad \left(\vec{\nu}_i - \frac{\vec{\varepsilon}}{4}\right)/\left(1 + \frac{l}{A \cdot b}\right) \geq \vec{\nu}_i - \frac{\vec{\varepsilon}}{2}$$

$$\text{iff} \quad \left(\vec{\nu}_i - \frac{\vec{\varepsilon}}{4}\right) \cdot A \cdot b \geq \left(\vec{\nu}_i - \frac{\vec{\varepsilon}}{2}\right) \cdot (A \cdot b + l)$$

$$\text{iff} \quad \frac{\vec{\varepsilon}}{2} \cdot A \cdot b \geq l \cdot \left(\vec{\nu}_i - \frac{\vec{\varepsilon}}{2}\right)$$

$$\text{iff} \quad \forall j \quad \frac{\varepsilon}{2} \cdot A \cdot b \geq l \cdot \left(\vec{\nu}_i[j] - \frac{\varepsilon}{2}\right)$$

$$\text{iff} \quad \forall j \quad A \geq \frac{l \cdot (2\vec{\nu}_i[j] - \varepsilon)}{\varepsilon \cdot b}$$

$$\text{if} \quad A \geq A_1 = \frac{l \cdot (2 \cdot W - \varepsilon)}{\varepsilon \cdot b}$$

where the first step follows from the inequality in Eq. 5, and the last step follows from $\frac{l \cdot (2 \cdot W - \varepsilon)}{\varepsilon \cdot b} \geq \frac{l \cdot (2\vec{\nu}_i[j] - \varepsilon)}{\varepsilon \cdot b}$, since, for an achievable vector $\vec{\nu}_i$, it must clearly be the case that $\vec{\nu}_i[j] \leq W$, the largest absolute value of any weight in the game.

Thanks to the two assumptions above, we can derive the following bound on $\vec{e}$:

$$\vec{e} \geq \frac{x_1 \cdot \vec{e}_{(b),1} + \cdots + x_k \cdot \vec{e}_{(b),k}}{1 + \frac{l}{A \cdot b}} - \frac{\vec{W} \cdot l}{A \cdot b + l}$$

$$\geq \frac{x_1 \cdot \vec{e}_{(b),1} + \cdots + x_k \cdot \vec{e}_{(b),k}}{1 + \frac{l}{A \cdot b}} - \frac{\vec{\varepsilon}}{2}$$

$$\geq x_1 \cdot \left(\vec{\nu}_1 - \frac{\vec{\varepsilon}}{2}\right) + \cdots + x_k \cdot \left(\vec{\nu}_k - \frac{\vec{\varepsilon}}{2}\right) - \frac{\vec{\varepsilon}}{2}$$

$$= x_1 \cdot \vec{\nu}_1 + \cdots + x_k \cdot \vec{\nu}_k - \vec{\varepsilon}$$

$$= \vec{\nu} - \vec{\varepsilon}$$

where the second inequality follows from (A), and the third inequality follows from (B). We also use the property that $\sum_{i=1}^{k} x_i = 1$ and $x_1 \cdot \vec{\nu}_1 + \cdots + x_k \cdot \vec{\nu}_k = \vec{\nu}$.

Now that we have a lower bound on the expected mean payoff over a single phase, we can show that the expected mean payoff of g_A over longer and longer prefixes of an infinite run (thus spanning many phases) converges to $\vec{e}$, from which the lemma follows. Notice that the expected mean payoff over m phases is simply $m \cdot \vec{e}$. Let $\vec{e}_n$ be the expected mean payoff of g_A in the first n steps. Since in n steps there are an expected number of $\left\lfloor \frac{n}{A \cdot b + l} \right\rfloor$ phases of expected length $A \cdot b + l$ each, we obtain

$$\vec{e}_n \geq \frac{\left\lfloor \frac{n}{A \cdot b + l} \right\rfloor \cdot (A \cdot b + l) \cdot \vec{e} + \left(n - \left\lfloor \frac{n}{A \cdot b + l} \right\rfloor \cdot (A \cdot b + l)\right) \cdot (-\vec{W})}{n}$$

$$\geq \frac{\left\lfloor \frac{n}{A \cdot b + l} \right\rfloor \cdot (A \cdot b + l) \cdot \vec{e} - (A \cdot b + l) \cdot \vec{W}}{n}$$

$$\geq \frac{(n - (A \cdot b + l)) \cdot \vec{e} - (A \cdot b + l) \cdot \vec{W}}{n}$$

and thus $\liminf_n \vec{e}_n = \vec{e}$. To conclude, take $A_\varepsilon := \max\{A_0, A_1\} = O(\frac{1}{\varepsilon})$, and the claim is satisfied for any $A \geq A_\varepsilon$. $\qquad \square$

Lemma 17. *For every $A \geq n$, $\mathcal{G}[g_A]$ is unichain.*

Proof. Each $\mathcal{G}_i[g_i]$ is recurrent by Lemma 14. By the definition of g_A, $\mathcal{G}[g_A]$ is obtained by staying in $\mathcal{G}_i[g_i]$ for a certain number of steps and then going to $\mathcal{G}_j[g_j]$ with $j = (i \mod k) + 1$ almost surely. If we stay in $\mathcal{G}_i[g_i]$ for at least n steps, then with positive probability we can visit every state therein. Thus, $\mathcal{G}[g_A]$ is unichain. $\qquad \square$

In the following, for every $\varepsilon > 0$, we denote by g_ε the strategy $g_{\max\{n, A_\varepsilon\}}$, where A_ε is the bound provided by Lemma 16.

Proof of Lemma 1. W.l.o.g. we assume that $\mathcal{G}$ reduces to a single end-component S. Let $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}^{+}(s_0)$. There exists $\vec{\nu}^* \in \mathsf{ExpSol}_{\mathcal{G}}(s_0)$ s.t. $\vec{\nu} < \vec{\nu}^*$. Let $\varepsilon = \frac{1}{2} \cdot \min_{i=1}^{d}(\vec{\nu}^*[i] - \vec{\nu}[i]) > 0$ be the half of the minimal difference between $\vec{\nu}^*$ and $\vec{\nu}$ in any component. Take $g := g_\varepsilon$. Then, $(\vec{\nu}^* - \vec{\varepsilon}) \in \mathsf{ExpSol}_{\mathcal{G}}(s_0, g)$ by Lemma 16. By the choice of ε, $\vec{\nu} < \vec{\nu}^* - \vec{\varepsilon}$, and thus $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}^{+}(s_0, g)$. Finally, $\mathcal{G}[g]$ is unichain by Lemma 17. $\qquad \square$

The next lemma will be used later to derive a bound on the number of steps K that strategy g_ε should be played for in order to have a mean payoff close to $\vec{\nu}$ with high probability.

Lemma 18. *For any $\varepsilon > 0$, there exist $K_0 \in \mathbb{N}$ and constants $a, b > 0$ s.t., for any $K \geq K_0$ and state s_0,*

$$\mathbb{P}^{\mathcal{G}}_{s_0, g_{\varepsilon/2}} \left[\exists (1 \leq j \leq d) \cdot |(\mathsf{MP}_K - \vec{\nu})[j]| \geq \varepsilon \right] \leq \mathcal{G}(K, \varepsilon/2) := 2^d \cdot a \cdot e^{-b \cdot K \cdot \varepsilon^2 / 4} .$$

Proof. Fix an error $\varepsilon > 0$, and let $\vec{\nu}^* = \mathbb{E}^{\mathcal{G}}_{s_0, g_{\varepsilon/2}}[\mathsf{MP}]$ the expected mean payoff when playing according to $g_{\varepsilon/2}$. By the choice of $g_{\varepsilon/2}$, $\vec{\nu} - \varepsilon/2 \leq \vec{\nu}^* \leq \vec{\nu}$. Thus, $|(\mathsf{MP}_K - \vec{\nu})[j]| \geq \varepsilon$ implies $|(\mathsf{MP}_K - \vec{\nu}^*)[j]| \geq \varepsilon/2$, and we have

$$\mathbb{P}^{\mathcal{G}}_{s_0, g_{\varepsilon/2}} \left[\exists (1 \leq j \leq d) \cdot |(\mathsf{MP}_K - \vec{\nu})[j]| \geq \varepsilon \right] \leq \mathbb{P}^{\mathcal{G}}_{s_0, g_{\varepsilon/2}} \left[\exists (1 \leq j \leq d) \cdot |(\mathsf{MP}_K - \vec{\nu}^*)[j]| \geq \varepsilon/2 \right] . \tag{6}$$

Since $\mathcal{G}[g_{\varepsilon/2}]$ is unichain by Lemma 17, we can apply Lemma 9, and, thus, there exist K_0 and constants $a, b > 0$ s.t., for every state s and $K \geq K_0$,

$$\mathbb{P}^{\mathcal{G}}_{s, g_{\varepsilon/2}} \left[\exists (1 \leq j \leq d) \cdot |\mathsf{MP}_K[j] - \vec{\nu}^*[j]| \geq \varepsilon/2 \right] \leq \mathcal{G}(K, \varepsilon/2) ,$$

from which the claim follows by Eq. 6. $\qquad\square$

APPENDIX B
BEYOND WORST-CASE SYNTHESIS

A. Finite-memory synthesis

In this section, we give full proofs for some statements from Sec. III-A. Fix a game $\mathcal{G}$ and worst-case threshold $\vec{\mu} := \vec{0}$.

Proposition 2. *Let f be a finite-memory strategy satisfying the worst-case threshold problem. The set of states visited infinitely often under f is almost surely a winning EC.*

Proof. By Proposition 1, the set of states visited infinitely often by f is an EC U almost surely. By contradiction, assume that U is not winning with some positive probability. Since f is finite-memory, $\mathcal{G}[f]$ is finite. Since U is visited with positive probability, there exists a reachable bottom strongly connected component B in $\mathcal{G}[f]$ which projected to $\mathcal{G}$ is a subset of U. Since U is not winning, there exists a play in B with mean payoff $\not\geq \vec{0}$. By prefix independence of the mean-payoff value function, there exists a play in $\mathcal{G}[f]$ with mean payoff $\not\geq \vec{0}$, contradicting that f is surely winning. $\qquad\square$

Lemma 3. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP, let s_0 be a state in a WEC W of $\mathcal{G}$, and let $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G} \downarrow W}(s_0)$ with $\vec{\nu} \geq \vec{0}$ be an expectation achievable by remaining inside W. There exists a randomized finite-memory strategy $h \in \Delta_F(\mathcal{G})$ s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}^+_{\mathcal{G} \downarrow W}(s_0, h)$ that also remains inside W.*

Remark 9. The statement of the lemma holds even with h a *pure* finite-memory strategy, by applying Remark 1 in the construction of strategy $g_{\frac{\varepsilon}{2}}^{exp}$ below.

The rest of this section is devoted to the proof of Lemma 3. To simplify the notation, we assume w.l.o.g. that the MDP $\mathcal{G}$ reduces to a single WEC $W = S$. Therefore, $\mathcal{G} = \mathcal{G} \downarrow W$. In the following, let

$$\mathsf{WCSol}_{\mathcal{G}}(s_0, f) = \{\vec{\mu} \in \mathbb{R}^d \mid \forall \pi \in \Omega^f_{s_0} \cdot \mathsf{MP}(\pi) \geq \vec{\mu}\} ,$$
$$\mathsf{WCSol}_{\mathcal{G}}(s_0) = \bigcup_{f \in \Delta(\mathcal{G})} \mathsf{WCSol}_{\mathcal{G}}(s_0, f) .$$

Recall that $\Omega^f_{s_0}$ is the set of plays originating from state s_0 which are consistent with strategy f. Since we are in a WEC, $\vec{0} \in \mathsf{WCSol}^+_{\mathcal{G}}(s_0)$, and thus $\mathsf{WCSol}_{\mathcal{G}}(s_0)$ contains a vector $\vec{\mu} > \vec{0}$. Moreover, since $\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G}}(s_0)$ with $\vec{\nu} \geq 0$, we can assume that $\vec{\nu} \in \mathsf{ExpSol}_{\mathcal{G}}(s_0)$ with $\vec{\nu} > 0$. W.l.o.g. we further assume $\vec{\mu} < \vec{\nu}$. We show that, for every $\delta > 0$ and $\varepsilon > 0$, there exists a randomized finite-memory strategy $h^{cmb}_{\delta, \varepsilon}$ satisfying the BWC threshold $\geq (\vec{\mu} - \vec{\delta}; \vec{\nu} - \vec{\varepsilon})$. The construction of $h^{cmb}_{\delta, \varepsilon}$ relies on the existence of the following two strategies:

- Since $\vec{\mu}$ is guaranteed in the worst case, by Lemma 2, for every $\delta > 0$ there exists a pure finite-memory strategy f^{wc}_δ s.t. $(\vec{\mu} - \vec{\delta}) \in \mathsf{WCSol}_{\mathcal{G}}(s, f^{wc}_\delta)$ for every state s in the WEC.
- Similarly, since $\vec{\nu}$ is achievable in expectation from state s_0, by Lemma 1, for every $\varepsilon > 0$ there exists a randomized finite-memory strategy g^{exp}_ε s.t. $(\vec{\nu} - \vec{\varepsilon}) \in \mathsf{ExpSol}_{\mathcal{G}}(s, g^{exp}_\varepsilon)$ for every state s in the WEC.

The strategy $h^{cmb}_{\delta, \varepsilon}$ is parameterised by two natural numbers $K, L \in \mathbb{N}$, and it alternates between $g^{exp}_{\frac{\varepsilon}{2}}$ and $f^{wc}_{\frac{\delta}{2}}$ over periods of length K and L, respectively:

(a) Play $g^{exp}_{\frac{\varepsilon}{2}}$ for $K > 0$ steps and record in $\overrightarrow{\mathsf{Sum}} \in \mathbb{Z}^d$ the current sum of the weights since the beginning of the period.

(b) If $\overrightarrow{\mathsf{Sum}} \geq (\vec{\mu} - \vec{\delta}) \cdot K$, then go to *(a)*. Otherwise, play $f^{wc}_{\frac{\delta}{2}}$ for $L > 0$ steps, and then go to *(a)*.

Every time a new period starts, the memory of the relevant strategy is reset. It remains to determine the values of the parameters K and L to reach the desired accuracy. The parameter K, which depends on ε, controls the probability that a period is of type *(a)* or *(a)+(b)*, and thus the quality of the approximation of the expectation objective: the larger the K, the higher the probability that a period is of type *(a)*, and thus the closer the expectation to $\vec{\nu}$. The parameter L (dependent on K and δ) controls the length of the recovery period *(a)+(b)*, and thus the larger the L, the closer the worst case to $\vec{\mu}$.

We show that we can always choose L s.t. the worst-case objective is satisfied.

Lemma 19. *For every $\delta, \varepsilon > 0$ and $K \in \mathbb{N}$, there exists $L \in \mathbb{N}$ s.t. $L = O(\frac{1}{\delta})$ and $(\vec{\mu} - \vec{\delta}) \in \mathsf{WCSol}_\mathcal{G}(s, h^{cmb}_{\delta,\varepsilon})$ for every state s in the EC.*

Proof. Let m be the product of the size of the memory of $f^{wc}_{\frac{\delta}{2}}$ and the number of states in $\mathcal{G}$, and let $\mu^* > 0$ be the smallest component of $\vec{\mu}$. W.l.o.g. we assume $\delta < \mu^*$, since $\mathsf{WCSol}_\mathcal{G}(s, h^{cmb}_{\delta,\varepsilon})$ is downward-closed. Below, we derive an expression for L for the worst-case objective $\geq \vec{\mu} - \vec{\delta}$ to be satisfied. Let π be any $h^{cmb}_{\delta,\varepsilon}$-consistent play. We decompose $\pi = \rho_0 \rho_1 \cdots$ according to periods *(a)* or *(a)+(b)*. If ρ_i has type *(a)*, then $\mathsf{TP}(\rho_i) \geq (\vec{\mu} - \vec{\delta}) \cdot K$ directly from the definition of periods of type *(a)*. If ρ_i has type *(a)+(b)*, then at the end of the *(a)* part (of length K) the sum of weights is at least $-K \cdot W$ in every component. (Recall that W is the largest absolute value of any weight in $\mathcal{G}$.) Moreover, during the following *(b)* part (of length L), we have that every time the same memory state of $h^{cmb}_{\delta,\varepsilon}$ and state of $\mathcal{G}$ repeats, the mean payoff is at least $\vec{\mu} - \frac{\vec{\delta}}{2}$, thus yielding a sum of weights which is at least $-m \cdot W + (L - m) \cdot (\mu^* - \frac{\delta}{2})$ in every component. Thus, for every component $0 \leq j < k$, we have

$$\mathsf{TP}(\rho_i)[j] \geq -K \cdot W - m \cdot W + (L - m) \cdot \left(\mu^* - \frac{\delta}{2} \right) .$$

In order to have $\mathsf{MP}(\pi) \geq \vec{\mu} - \vec{\delta}$, it suffices to have $\mathsf{MP}(\rho_i) \geq \vec{\mu} - \vec{\delta}$ for each period i since periods are of uniformly bounded length, and thus $\mathsf{TP}(\rho_i) \geq (\vec{\mu} - \vec{\delta}) \cdot (K + L)$, since this length is at most $K + L$. It is easy to see that the following choice for L satisfies the constraint above:

$$L := \left\lceil \frac{2 \cdot K \cdot (W + \mu^* - \delta) + m \cdot (2 \cdot W + 2 \cdot \mu^* - \delta)}{\delta} \right\rceil \tag{7}$$

$\square$

In the following, we consider L as fixed by Eq. 7. We show that one can always choose K s.t. the expectation objective is satisfied. We crucially use the fact that L is linear in K, and that the probability of periods of type *(a)+(b)* can be made negligible for large K.

Lemma 20. *For sufficiently small $\delta, \varepsilon > 0$, there exists $K \in \mathbb{N}$ s.t. $K = O(\frac{1}{\varepsilon})$, and $(\vec{\nu} - \vec{\varepsilon}) \in \mathsf{ExpSol}_\mathcal{G}(s, h^{cmb}_{\delta,\varepsilon})$ for every state s in the EC.*

Proof. Let $\vec{E}(K, s)$ be the expected mean payoff vector in the MDP $\mathcal{G}$ when Controller is playing according to $h^{cmb}_{\delta,\varepsilon}$, starting from state s. We prove that there exists a $K = O(\frac{1}{\varepsilon})$ s.t.

$$\vec{E}(K, s) \geq \vec{\nu} - \vec{\varepsilon} .$$

Let $\vec{E}_{(a)}(K, s)$ and $\vec{E}_{(a)+(b)}(K, s)$ be the expected mean payoff of periods of type *(a)* and *(a)+(b)*, respectively. Let $p(K)$ be the probability of having a period of type *(a)+(b)*. The expected length of a period is $(1 - p(K)) \cdot K + p(K) \cdot (K + L)$. Similarly, the expected total payoff of a period is $(1 - p(K)) \cdot \vec{E}_{(a)}(K, s) \cdot K + p(K) \cdot \vec{E}_{(a)+(b)}(K, s) \cdot (K + L)$. We thus obtain the following expression on the right for the expected mean payoff over one period, where the equality to the expected mean payoff over the entire play is easy to show:

$$\vec{E}(K, s) = \frac{(1 - p(K)) \cdot \vec{E}_{(a)}(K, s) \cdot K + p(K) \cdot \vec{E}_{(a)+(b)}(K, s) \cdot (K + L)}{(1 - p(K)) \cdot K + p(K) \cdot (K + L)}$$

By dividing by $(1 - p(K)) \cdot K$, we obtain the following inequality:

$$\vec{E}(K, s) = \frac{\vec{E}_{(a)}(K, s) + \frac{p(K)}{1-p(K)} \cdot \vec{E}_{(a)+(b)}(K, s) \cdot \frac{K+L}{K}}{1 + \frac{p(K)}{1-p(K)} \cdot \frac{K+L}{K}} \geq \vec{\nu} - \vec{\varepsilon}$$

Since $\vec{E}_{(a)+(b)}(K, s) \geq \vec{\mu} - \vec{\delta}$ by the choice of L (cf. the proof of Lemma 19), and $\vec{\mu} - \vec{\delta} > \vec{0}$ for sufficiently small δ, it suffices to find K s.t.

$$\frac{\vec{E}_{(a)}(K, s)}{1 + \frac{p(K)}{1-p(K)} \cdot \frac{K+L}{K}} \geq \vec{\nu} - \vec{\varepsilon} \tag{8}$$

From a qualitative point of view, when taking the limit for $K \to \infty$ in Eq. 8,

- $\frac{K+L}{K}$ tends to a constant, since L is linear in K, and
- $p(K)$ tends to 0.

Thus, $\lim_{K \to \infty} \vec{E}(K, s) = \lim_{K \to \infty} \vec{E}_{(a)}(K, s) \geq \vec{\nu} - \frac{\vec{\varepsilon}}{2}$, where the inequality follows from the fact in periods of type *(a)* we are playing according to $g_{\frac{\vec{\varepsilon}}{2}}^{exp}$, which guarantees expectation $\geq \vec{\nu} - \frac{\vec{\varepsilon}}{2}$ in the long run. Consequently, for every $\varepsilon > 0$, there exists a K s.t. $\vec{E}(K, s) \geq \vec{\nu} - \vec{\varepsilon}$, as required.

The rest of the proof shows that K can be taken to be $O(\frac{1}{\varepsilon})$. We proceed by an Hoeffding-style analysis. The following assumption can be satisfied for sufficiently small $\delta, \varepsilon > 0$, since $\vec{\mu} < \vec{\nu}$ was assumed throughout this section.

Assumption 3. $\vec{\nu} - \vec{\varepsilon} \geq \vec{\mu} - \vec{\delta} \geq \vec{0}$.

Lemma 18 provides a constant K_0 and an upper bound $\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)$ on the probability that the mean payoff deviates from $\vec{\nu}$ by more than $\frac{\varepsilon}{2}$ in any component when playing according to $g_{\frac{\vec{\varepsilon}}{2}}^{exp}$ for every $K \geq K_0$. Thus, $1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)$ is a lower bound on the probability that the mean payoff in any component deviates from $\vec{\nu}$ by less than $\frac{\varepsilon}{2}$.

Assumption 4. *Let $K \geq K_0$.*

We get the following lower bound on the expected mean payoff $\vec{E}_{(a)}(K, s)$ of a period of type *(a)*:

$$\vec{E}_{(a)}(K, s) \geq (1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)) \cdot \left(\vec{\nu} - \frac{\vec{\varepsilon}}{2}\right) + \mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \cdot \vec{\nu}_0$$

where $\vec{\nu}_0 \geq \vec{0}$ is a lower bound on the mean payoff of any period of type *(a)*, Therefore, we obtain the following simpler bound:

$$\vec{E}_{(a)}(K, s) \geq (1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)) \cdot \left(\vec{\nu} - \frac{\vec{\varepsilon}}{2}\right) \tag{9}$$

Moreover, we can use Lemma 18 also to provide a bound on $p(K)$, the probability of a period of type *(a)+(b)*. If a period is of type *(a)+(b)*, then there exists a component $1 \leq j \leq d$ s.t. $\mathsf{TP}_K[j] \leq 0$, and thus $\mathsf{MP}_K[j] \leq 0$. Since the strategy $g_{\frac{\vec{\varepsilon}}{2}}^{exp}$ achieves expectation $\geq \vec{\nu} - \frac{\vec{\varepsilon}}{2} \geq \vec{0}$, we have $|(\mathsf{MP}_K - \vec{\nu})[j]| \geq \frac{\varepsilon}{2}$ for some $1 \leq j \leq d$. The latter event happens with probability at most $\mathcal{G}\left(K, \frac{\vec{\varepsilon}}{4}\right)$ by Lemma 18. We thus obtain the following bound:

$$p(K) \leq \mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \tag{10}$$

By using Eq. 9 and 10 in Eq. 8, it suffices to find a K s.t.

$$\frac{(1 - \mathcal{G}\left(K, \frac{\vec{\varepsilon}}{4}\right)) \cdot (\vec{\nu} - \frac{\vec{\varepsilon}}{2})}{1 + \frac{\mathcal{G}\left(K, \frac{\vec{\varepsilon}}{4}\right)}{1 - \mathcal{G}\left(K, \frac{\vec{\varepsilon}}{4}\right)} \cdot \frac{K+L}{K}} \geq \vec{\nu} - \vec{\varepsilon}$$

Let γ be an upper bound on $\frac{\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)}{1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)} \cdot \frac{K+L}{K}$. Then, it suffices to show

$$\forall (1 \leq j \leq d) \qquad \frac{(1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)) \cdot (\vec{\nu}[j] - \frac{\varepsilon}{2})}{1 + \gamma} \geq \vec{\nu}[j] - \varepsilon$$

$$\text{if} \quad \forall (1 \leq j \leq d) \qquad 1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \geq \frac{(\vec{\nu}[j] - \varepsilon)(1 + \gamma)}{\vec{\nu}[j] - \frac{\varepsilon}{2}}$$

$$\text{if} \quad \forall (1 \leq j \leq d) \qquad 1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \geq \frac{\vec{\nu}[j] - \varepsilon + \gamma \cdot (\vec{\nu}[j] - \varepsilon)}{\vec{\nu}[j] - \frac{\varepsilon}{2}}$$

$$\text{if} \quad \forall (1 \leq j \leq d) \qquad \mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \leq \frac{\frac{\varepsilon}{2} - \gamma \cdot (\vec{\nu}[j] - \varepsilon)}{\vec{\nu}[j] - \frac{\varepsilon}{2}} \tag{11}$$

Let $\nu^* = \max\{\vec{\nu}[1], \ldots, \vec{\nu}[d]\}$ be the maximal component of $\vec{\nu}$. We show that with the following value of γ we can satisfy all objectives.

Assumption 5. $\gamma := \frac{1}{4} \cdot \frac{\varepsilon}{\nu^* - \varepsilon}$.

We now plug in the value for γ into Eq. 11, to obtain $\mathcal{G}\left(K, \varepsilon/4\right) \leq \frac{\frac{\varepsilon}{2} - \frac{1}{4} \cdot \frac{\varepsilon}{\nu^* - \varepsilon} \cdot (\nu^* - \varepsilon)}{\nu^* - \frac{\varepsilon}{2}}$, yielding the following bound on $\mathcal{G}\left(K, \varepsilon/4\right)$:

$$\mathcal{G}\left(K, \varepsilon/4\right) \leq \frac{1}{2} \cdot \frac{\varepsilon}{2\nu^* - \varepsilon} \tag{12}$$

Before solving the inequality above for K, we go back one step discuss the constraint derived from the definition of γ. Recall that γ should satisfy $\frac{\mathcal{G}(K,\varepsilon/4)}{1-\mathcal{G}(K,\varepsilon/4)} \cdot \frac{K+L}{K} \leq \gamma$. By replacing γ with its definition $\gamma := \frac{1}{4} \cdot \frac{\varepsilon}{\nu^*-\varepsilon}$ in the latter inequality, we obtain the following inequality:

$$\frac{\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)}{1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)} \cdot \frac{K+L}{K} \leq \frac{1}{4} \cdot \frac{\varepsilon}{\nu^* - \varepsilon} \tag{13}$$

By the definition, $L = \left\lceil \frac{2 \cdot K \cdot (W + \mu^* - \delta) + m \cdot (2 \cdot W + 2 \cdot \mu^* - \delta)}{\delta} \right\rceil$, and thus $L \leq \left\lceil \frac{2 \cdot K \cdot (W + \mu^*) + m \cdot (2 \cdot W + 2 \cdot \mu^*)}{\delta} \right\rceil \leq \frac{2 \cdot (W + \mu^*) \cdot (K + m)}{\delta} + 1$.

Thus, $\frac{K+L}{K} \leq \frac{K + \frac{2 \cdot (W+\mu^*) \cdot (K+m)}{\delta} + 1}{K} = 1 + \frac{2(W+\mu^*)}{\delta} + \frac{1}{K} \cdot \frac{2m(W+\mu^*)+\delta}{\delta}$. In particular, for $K \geq \frac{2m(W+\mu^*)+\delta}{\delta}$, we have $\frac{K+L}{K} \leq 2 + \frac{2(W+\mu^*)}{\delta}$.

Assumption 6. $K \geq K_1 := \max\{K_0, \frac{2m(W+\mu^*)+\delta}{\delta}\}$.

Thus, for $K \geq K_1$, it suffices to satisfy the following inequality:

$$\frac{\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)}{1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)} \cdot \left(2 + \frac{2(W+\mu^*)}{\delta}\right) \leq \frac{1}{4} \cdot \frac{\varepsilon}{\nu^* - \varepsilon} \tag{14}$$

We perform the following algebraic manipulations:

$$\frac{\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)}{1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)} \cdot \frac{2}{\delta} \cdot (W + \mu^* + \delta) \leq \frac{1}{4} \cdot \frac{\varepsilon}{\nu^* - \varepsilon}$$

if
$$\mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \leq \frac{1}{8} \cdot \frac{\delta}{W + \mu^* + \delta} \cdot \frac{\varepsilon}{\nu^* - \varepsilon} \cdot \left(1 - \mathcal{G}\left(K, \frac{\varepsilon}{4}\right)\right)$$

if
$$\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)\left(1 + \frac{1}{8} \cdot \frac{\delta}{W + \mu^* + \delta} \cdot \frac{\varepsilon}{\nu^* - \varepsilon}\right) \leq \frac{1}{8} \cdot \frac{\delta}{W + \mu^* + \delta} \cdot \frac{\varepsilon}{\nu^* - \varepsilon}$$

if
$$\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)\left(8(W + \mu^* + \delta)(\nu^* - \varepsilon) + \delta\varepsilon\right) \leq \delta\varepsilon$$

if
$$\mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \leq \frac{\delta\varepsilon}{8(W + \mu^* + \delta)(\nu^* - \varepsilon) + \delta\varepsilon}$$

if
$$\mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \leq \frac{\delta\varepsilon}{(8\mu^* + 8W + 8\delta) \cdot \nu^* - (8\mu^* + 8W + 7\delta) \cdot \varepsilon}$$

if
$$\mathcal{G}\left(K, \frac{\varepsilon}{4}\right) \leq \frac{\varepsilon}{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon} \tag{15}$$

We now compare the bounds on $\mathcal{G}\left(K, \frac{\varepsilon}{4}\right)$ given by Eq. 12 and 15, and we show that, for sufficiently small ε, the latter bound implies the former. Therefore, we seek for ε s.t.

$$\frac{\varepsilon}{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon} \leq \frac{1}{2} \cdot \frac{\varepsilon}{2\nu^* - \varepsilon}$$

if
$$\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon \geq 4 \cdot \nu^* - 2 \cdot \varepsilon$$

if
$$\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 4\right) \cdot \nu^* \geq \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 5\right) \cdot \varepsilon$$

if
$$\varepsilon \leq \frac{8(\mu^* + W) + 4\delta}{8(\mu^* + W) + 5\delta} \cdot \nu^*$$

Assumption 7. $\varepsilon \leq \varepsilon_0 := \frac{8(\mu^* + W) + 4\delta}{8(\mu^* + W) + 5\delta} \cdot \nu^*$.

Therefore, for sufficiently small $\varepsilon \leq \varepsilon_0 < \nu^*$ it suffices to solve Eq. 15. Recall that $\mathcal{G}\left(K, \frac{\varepsilon}{4}\right) = a \cdot 2^d \cdot e^{-b \cdot K \cdot \varepsilon^2/16}$. Thus, we seek for K s.t.

$$a \cdot 2^d \cdot e^{-b \cdot K \cdot \varepsilon^2/16} \leq \frac{\varepsilon}{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon}$$

if
$$e^{-b \cdot K \cdot \varepsilon^2/16} \leq \frac{1}{a \cdot 2^d} \cdot \frac{\varepsilon}{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon}$$

if
$$-b \cdot K \cdot \varepsilon^2/16 \leq \ln\left(\frac{1}{a \cdot 2^d} \cdot \frac{\varepsilon}{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon}\right)$$

if
$$K \geq \frac{16}{b \cdot \varepsilon^2} \cdot \ln\left(a \cdot 2^d \cdot \frac{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^* - \left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 7\right) \cdot \varepsilon}{\varepsilon}\right)$$

if
$$K \geq \frac{16}{b \cdot \varepsilon^2} \cdot \ln\left(a \cdot 2^d \cdot \frac{\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) \cdot \nu^*}{\varepsilon}\right)$$

if
$$K \geq \frac{16}{b \cdot \varepsilon^2} \cdot \left(\ln a + d\ln 2 + \ln\left(\frac{8\mu^*}{\delta} + \frac{8W}{\delta} + 8\right) + \ln \nu^* - \ln \varepsilon\right) =: K_2 \qquad (16)$$

The constant K_2 is of the form $\frac{\alpha}{\varepsilon^2}$, with α linear in ε and d (recall that a is exponential in ε), and polynomial in the characteristics of the Markov chain. Thus, $K = O(\frac{1}{\varepsilon})$. $\qquad \square$

B. Infinite-memory synthesis

1) Inside an EC: We complete the proof of Lemma 6 by showing the two claims. The first claim relies on Lemma 3.9 in [19], while the second claim relies on Lemma 18.

Claim 1. *There are rational constants a and b with $b < 1$ s.t., for every $i \geq 1$ and for sufficiently large K, $p_{i,K} \leq a \cdot b^{K \cdot i}$. Note that a and b do not depend neither on K, nor on i.*

Proof. Recall that $p_{i,K}$ is, by definition, the probability that the total payoff goes below $\vec{N}_i$ in any component during phase i. Since at the beginning of phase i the total payoff is $\mathsf{TP}_{K \cdot i} > 2 \cdot \vec{N}_i$ by definition, $p_{i,K}$ is upper bounded by the probability that the total payoff decreases by at least $\vec{N}_i$ in any component:

$$p_{i,K} \leq \mathbb{P}_{s_0}^{\mathcal{G}[f_K]}\left[\exists(K \cdot i \leq h < K \cdot (i+1)) \cdot \mathsf{TP}_h - \mathsf{TP}_{K \cdot i} \not\succ -\vec{N}_i\right]$$

For any state s in the EC, and payoff threshold $N \in \mathbb{N}$, let $p_{s,N,K}$ be the probability that the total payoff goes below $-N$ in any component when starting from $\vec{0}$ within at most K steps:

$$p_{s,N,K} = \mathbb{P}_s^{\mathcal{G}[f_K]}\left[\exists(0 \leq h < K) \cdot \exists(1 \leq j \leq d) \cdot \mathsf{TP}_h[j] \leq -N\right]$$

Thus, $p_{i,K} \leq \min_{s \in S} p_{s, \min \vec{N}_i, K}$. We prove the claim thanks to the following lemma. (It follows from Lemma 3.9 in [19]. For completeness, we provide a proof below.)

Lemma 21. *There exists a rational constant $c < 1$ and an integer $N^* \geq 0$ s.t., for every $N \geq N^*$ and starting state s in $\mathcal{G}[f_K]$ s.t. $\mathbb{P}_s^{\mathcal{G}[f_K]}\left[\mathsf{MP} \geq \vec{0}\right] = 1$,*

$$p_{s,N,K} \leq 2^d \cdot \frac{c^N}{1-c}$$

By Lemma 1 we can assume that when playing according to f^{exp} we obtain a Markov chain $\mathcal{G}[f^{exp}]$ which is *unichain*. Since f^{exp} has strictly positive expected mean payoff and $\mathcal{G}[f^{exp}]$ is unichain, the mean payoff is almost surely strictly positive, and thus we can apply Lemma 21 to the Markov chain $\mathcal{G}[f^{exp}]$ and obtain constants N^* and c s.t. for every $N \geq N^*$, $K \geq 0$, and state s, $p_{s,N,K} \leq 2^d \frac{c^N}{1-c}$. Let $\nu^* = \min_{1 \leq j \leq d} \vec{v}[j]$, and let K^* be large enough s.t., for any $K \geq K^*$, $N_i^{min} \geq N^*$. (For example, take $K^* = \frac{2N^*}{\nu^*}$.) Assume we are in any phase $i \geq 1$. Since $N_i^{min} \geq N^*$ by the choice of K^*, by Lemma 21 we obtain, for every $K \geq K^*$, $p_{s,N_i^{min},K} \leq 2^d \cdot \frac{c^{N_i^{min}}}{1-c} = 2^d \cdot \frac{c^{\frac{\nu^* \cdot i \cdot K}{2}}}{1-c}$. Since $p_{i,K} \leq \min_{s \in S} p_{s, \min \vec{N}_i, K}$, we have $p_{i,K} \leq 2^d \frac{c^{\frac{\nu^* \cdot i \cdot K}{2}}}{1-c}$. The claim follows by taking $a = 2^d/(1-c)$ and $b = c^{\frac{\nu^*}{2}}$, which is < 1 since c is. $\qquad \square$

Proof of Lemma 21. We first prove the lemma in the unidimensional case. Let T be the set of states from where the total payoff is almost surely $-\infty$, and let T^* be the set of states which can reach T with positive probability. Lemma 3.9 in [19]

when applied to the Markov chain $\mathcal{G}[f_K]$ (in general, it applies to Markov decision processes), guarantees that there exists a rational constant $c < 1$ and an integer $N^* \geq 0$ s.t. for every $N \geq N^*$

$$\mathbb{P}_s^{\mathcal{G}[f_K]}\left[\exists i \in \mathbb{N} \cdot \mathsf{TP}_i \leq -N \text{ and } T^* \text{ is not visited }\right] \leq \frac{c^N}{1-c}$$

Since we assume $\mathbb{P}_s^{\mathcal{G}[f_K]}[\mathsf{MP} \geq 0] = 1$, by Lemma 12 the probability to have a total payoff equal to $-\infty$ is 0, and thus T is empty. Therefore, the probability on the left above is just $\mathbb{P}_s^{\mathcal{G}[f_K]}[\exists i \in \mathbb{N} \cdot \mathsf{TP}_i \leq -N]$, which is the definition of $p_{s,N,K}$, and thus

$$p_{s,N,K} \leq \frac{c^N}{1-c}$$

Now let $d > 1$, and, for a component $1 \leq j \leq d$, let $p_{s,N,K}^j$ be the probability that we have a drop by $-N$ in the total payoff in component j, i.e.,

$$p_{s,N,K}^j = \mathbb{P}_s^{\mathcal{G}[f_K]}[\exists i \in \mathbb{N} \cdot \mathsf{TP}_i[j] \leq -N]$$

By looking at the Markov chain $\mathcal{G}[f_K]$ projected to component j, we can apply the result from the first part and obtain that, for every fixed j, there exist $c_j < 1$ and $N_j^* \geq 0$ s.t., for every $N \geq N_j^*$,

$$p_{s,N,K}^j \leq \frac{c_j^N}{1-c_j}$$

By taking $N^* = \max_{1 \leq j \leq d} N_j^*$ and letting c be the c_j maximizing $\frac{c_j}{1-c_j}$, we have, for every $N \geq N^*$,

$$p_{s,N,K}^j \leq \frac{c^N}{1-c}$$

By definition, $1 - p_{s,N,K}^j$ is the probability that the total payoff never goes below $-N$ in component j, and thus $\Pi_{j=1}^d(1-p_{s,N,K}^j)$ is a lower bound on the probability that the total payoff never goes below $-N$ in *any component*. Therefore $p_{s,N,K} \leq 1-\Pi_{j=1}^d(1-p_{s,N,K}^j) \leq 1-(1-\frac{c^N}{1-c})^d$. By a simple calculation (cf. the proof of Lemma 9), we derive $(1-\frac{c^N}{1-c})^d \geq 1-2^d \cdot \frac{c^N}{1-c}$, and thus

$$p_{s,N,K} \leq 1 - \left(1 - 2^d \cdot \frac{c^N}{1-c}\right) = 2^d \cdot \frac{c^N}{1-c}$$

as required. $\qquad\square$

Claim 2. *There exist rational constants a and b with $b < 1$ s.t., for every $i \geq 0$ and sufficiently large K, $q_{i,K} \leq a \cdot b^{K \cdot (i+1)}$. Note that a and b do not depend neither on K, nor on i.*

Proof. Recall the definition of $q_{i,K} = \mathbb{P}_{s_0}^{\mathcal{G}[f_K]}\left[\mathsf{TP}_{K \cdot (i+1)} \not\geq 2 \cdot \vec{N}_{i+1}\right]$. By the definition of $\vec{N}_{i+1} = \frac{\vec{v}(i+1)K}{2}$,

$$q_{i,K} = \mathbb{P}_{s_0}^{\mathcal{G}[f_K]}\left[\mathsf{MP}_{K \cdot (i+1)} \not\geq \vec{v}\right]$$

Let $\vec{v}^*$ be the expected mean payoff of f^{exp}, and let $\delta = \min_{1 \leq j \leq d}(\vec{v}^*[j] - \vec{v}[j]) > 0$. Then, $q_{i,K}$ is upper bounded by the probability that the mean payoff deviates from its expected value $\vec{v}^*$ by more than δ in any component. By Lemma 18, this is upper bounded by $a'2^d e^{-b'(i+1)K\delta^2}$ for sufficiently large $(i+1)K$, and thus for sufficiently large K. Take $a = a' \cdot 2^d$ and $b = e^{-b\delta^2} < 1$. $\qquad\square$

C. Inside a general MDPs

In this section, we prove the correctness of the reduction of the BWC problem to solving system T'.

Lemma 7. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP, let $\vec{v} \geq \vec{0}$, and let $s_0 \in S$. There exists a (possibly infinite-memory) strategy h s.t. $(\vec{0}; \vec{v}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h)$, if, and only if, the system T' has a non-negative solution.*

The following proposition is the analogous of Proposition 3 where we consider MEC (instead of MWEC) and arbitrary strategies (instead of finite-memory ones). As before, it rests on the assumption that all states of $\mathcal{G}$ are reachable from s_0.

Proposition 6. *Let $\mathcal{G}$ be a pruned multidimensional mean-payoff MDP. If there exists a strategy h s.t. $(\vec{0}; \vec{v}) \in \mathsf{BWCSol}_{\mathcal{G}}^+(s_0, h)$, then there exists a strategy h^* with the same property, and such that, for every MEC U, the set of states visited infinitely often by h^* is a subset of U with positive probability.*

The following proposition (and its proof) is analogous to Proposition 4, where MWECs are replaced by MECs.

Proposition 7. *If T' has a non-negative solution, then there exists a (finite-memory) strategy $\hat{h}$ s.t.*

1) *$\vec{\nu} \in \mathsf{ExpSol}^+_{\mathcal{G}}(s_0, \hat{h})$.*
2) *For every MEC U, there is a probability $y^*_U > 0$ s.t. the set of states visited infinitely often by $\hat{h}$ is inside U with probability y^*_U.*
3) *The set of states visited infinitely often by $\hat{h}$ is almost surely an EC. Consequently, $\sum_{MEC\ U} y^*_U = 1$.*
4) *Once inside a MEC U, $\hat{h}$ achieves expected mean payoff $\vec{\nu}_U > \vec{0}$.*
5) *$\sum_{MEC\ U} y^*_U \cdot \vec{\nu}_U > \vec{\nu}$.*

Proof of Lemma 7. The proof is similar to finite-memory case in Lemma 4. We sketch here the crucial differences.

For one direction, let h be a strategy s.t. $(\vec{0}; \vec{\nu}) \in \mathsf{BWCSol}^+_{\mathcal{G}}(s_0, h)$. By Proposition 6, there exists a strategy h^* that additionally visits each MEC infinitely often with positive probability. By the construction in Proposition 4.4 of [9] applied to strategy h^*, we obtain a solution to system T'. Equations (A1), (A1'), (A2), (C1), (C1'), and (C2) are shown to be satisfied in the proof of Proposition 4.4. Eq. (B-bis) is satisfied since, by Proposition 1, h^* is eventually trapped in an EC almost surely (not necessarily a WEC). Finally, Eq. (C3-bis) is satisfied: By construction, h^* visits each MEC infinitely often with positive probability. For every MWEC U there exist $s, t \in U$ s.t. $x^*_{st} > 0$. Since h^* is winning for the worst-case, it achieves an expected mean payoff $> \vec{0}$ in U, and thus Eq. (C3-bis) is satisfied.

For the other direction, we use Proposition 7 to obtain strategy $\hat{h}$, and then we proceed as in the second part of the proof of Lemma 4 by replacing MWEC with MEC, and by using Lemma 5 instead of Lemma 3 in the construction of strategies h_U's. $\qquad\square$

REFERENCES FOR THE APPENDIX

[17] M. Tracol, "Fast convergence to state-action frequency polytopes for mdps," *Oper. Res. Lett.*, vol. 37, no. 2, pp. 123–126, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.orl.2008.12.003
[18] A. Schrijver, *Theory of linear and integer programming*. Wiley Publishing, Inc., 1986.
[19] T. Brázdil, V. Brožek, K. Etessami, and A. Kučera, "Approximating the termination value of one-counter MDPs and stochastic games," *Information and Computation*, vol. 222, no. 0, pp. 121–138, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0890540112001599