Solving Robust POMDPs with ω-regular Objectives via Partially Observable Stochastic Games

Durgam Latha, Dion Reji, S. Akshay
Indian Institute of Technology Bombay
Mumbai, India

Đorđe Žikelić
Nanyang Technological University
Singapore, Singapore

Shankaranarayanan Krishna
Indian Institute of Technology Bombay
Mumbai, India

Abstract

Robust POMDPs (RPOMDPs) generalize classical POMDPs to the setting where exact transition probabilities are not known – rather, they are only known to belong to some uncertainty set of values. In this work, we study the problem of solving RPOMDPs with general ω-regular objectives, which subsume a broad class of objectives such as reachability, safety, and linear temporal logic (LTL) objectives. We show that, for (s, a)-rectangular RPOMDPs with polytopic uncertainty sets, the problem of solving RPOMDPs under ω-regular objectives can be reduced to solving partially observable stochastic games (POSGs) under ω-regular objectives. Moreover, we show for the first time that reductions can be constructed in both directions, establishing the semantic equivalence between (s, a)-rectangular RPOMDPs with polytopic uncertainty sets and POSGs. This allows us to derive a range of new computational complexity results, including both upper and lower complexity bounds, on solving RPOMDPs with different ω-regular objectives. As a corollary, we also derive new computational complexity results for RMDPs.

1 Introduction

Partially observable Markov decision processes (POMDPs) are a standard model for sequential decision making under uncertainty and imperfect state observability [23]. The problem of solving a POMDP is concerned with computing a strategy (or policy) that maximizes the expected payoff or the probability of satisfying some objective. However, classical algorithms assume that transition probabilities are exactly known, which is not always a realistic assumption. In practice, transition probabilities are usually inferred from data and as such their estimates come with a level of uncertainty [8]. In the fully observable setting, this challenge is addressed through robust MDPs (RMDPs) [30, 22], which generalize classical MDPs by assuming that transition probabilities belong to some *uncertainty sets* of possible values, rather than being exactly given. Robust POMDPs (RPOMDPs) [31] provide a robust generalization of POMDPs.

Recent years have seen significant advances in algorithmic approaches to solving RMDPs and RPOMDPs [30, 22, 46, 24, 38, 14, 20, 1, 40, 17, 16, 42, 8, 28, 31, 6, 36, 29, 11, 12, 13, 26]. However, reward objectives do not consider *logical correctness* properties that are particularly important in safety-critical applications. For instance, in self-driving cars, the reach-avoid correctness property is concerned with reaching some target region while avoiding the unsafe obstacles [37]. The stability correctness property requires the agent to eventually reach and stay within some target region ad infinitum [25]. These are both classical examples of logical correctness properties in control and robotics applications, which are not defined in terms of reward objectives. In formal methods, logical correctness properties are formally defined via logics such as linear temporal

logic (LTL) [32], or more generally ω-regular specifications [10], which subsumes most logical correctness properties appearing in control and robotics applications [27]. While recent years have seen significant advances in solving RMDPs and RPOMDPs with reward objectives, the literature on logical correctness objectives remains limited. Even the literature on RMDPs with ω-regular objectives either makes significant assumptions on uncertainty sets such as interval MDPs [9, 34], or is restricted to probability 1 or limit-sure satisfaction of ω-regular specifications [2].

Our Contributions. In this work, we consider the problem of solving RMDPs and RPOMDPs with ω-regular objectives, with a focus on computational complexity aspects. As common in the literature on solving R(PO)MDPs, we consider RPOMDPs with (s, a)-rectangular and polytopic uncertainty sets. These assumptions are quite general and subsume types of uncertainty sets considered in the existing literature, such as intervals [15] or L^1-uncertainty sets [20]. Our approach is based on the following fundamental novel result – we formally prove that the problem of solving RPOMDPs with ω-regular objectives is *equivalent* to the problem of solving partially observable stochastic games (POSGs) [19] with ω-regular objectives. That is, we show that the two problems can be reduced to each other in polynomial-time and thus any known algorithm and computational complexity result for solving one problem readily applies to the other problem. The key novelty of our result lies in establishing the equivalence of the two models. Previous work has shown that, under reward objectives, the problem of solving R(PO)MDPs can be reduced to the problem of solving (PO)SGs [8, 18, 5]. However, in this work, we consider ω-regular objectives rather than reward objectives. Moreover, we show that the two models can be reduced to each other, hence are equivalent, and the reduction from (PO)SGs to R(PO)MDPs turns out to be much more technically challenging. To the best of our knowledge, this is the first time that a stochastic game model has been formally polynomial-time reduced to a robust MDP model.

Since the problem of solving POSGs under ω-regular objectives has received more attention [7] compared to RPOMDPs, our equivalence result allows us to derive a plethora of new computational complexity results on solving RPOMDPs that were previously unknown. Our results are summarized in Table 1. We derive results for several variants of the ω-regular objective satisfaction problem: (1) Sure Winning, where the objective needs to be satisfied for every RPOMDP run, (2) Almost Sure Winning, where the objective needs to be satisfied with probability 1, (3) Limit Sure Winning, where the objective needs to be satisfied with probability $1 - \epsilon$ for every $\epsilon > 0$, and (4) Quantitative Winning, where the objective needs to be satisfied with at least some given probability $p \in [0, 1]$. Moreover, we derive complexity bounds both under the one-sided partial observability setting (that we call 1s-RPOMDPs) where only the agent has partial observability but the environment has full observability of the state, and the two-sided partial observability setting (which we refer to as RPOMDPs) where both the agent and the environment only have partial observability of the state. In addition, as a direct corollary of our equivalence result under full observability, we derive new complexity bounds for RMDPs with ω-regular objectives under the Quantitative Winning setting that were previously unknown, see Table 1. Thus, our work significantly advances the state of the art in solving R(PO)MDPs under ω-regular objectives and provides a complete picture of the complexity landscape for different variants of the problem. Our contributions can be summarized as follows:

1. **Equivalence of RPOMDPs and POSGs.** We formally prove, for the first time, that the problems of solving RPOMDPs and POSGs under ω-regular objectives are *equivalent*, i.e. polynomial-time reducible to each other. No previous work has considered reduction from a stochastic game model to a robust MDP model.
2. **New complexity results for R(PO)MDPs.** We derive new computational complexity results for RPOMDPs under ω-regular objectives, see Table 1. As a corollary, we also derive new complexity results for RMDPs.

Related Work. Recent years have seen significant advances in algorithmic approaches to solving RMDPs under reward objectives, with efficient algorithms being proposed for expected discounted-sum reward and long-run average reward, as discussed above. These works assume knowledge of the RMDP model and provide guarantees on the correctness of their results. In addition, reinforcement learning algorithms for solving RMDPs in a model-free setting but without guarantees on the correctness of results have also been proposed [33, 39, 44, 45, 43]. Algorithmic approaches for solving RPOMDPs under reward objectives were considered in [31, 6, 36, 29, 11, 12, 13, 26]. See [35] for a recent survey. We study R(PO)MDPs under ω-regular objectives, which have received much less attention. To the best of our knowledge, no existing work considers solving RPOMDPs with ω-regular objectives. Literature on RMDPs with ω-regular objectives either makes significant assumptions on

	Sure Winning (Table 1)			**Almost Sure Winning**(Table 3)		
	RMDP	**1s RPOMDP**	**RPOMDP**	**RMDP**	**1s RPOMDP**	**RPOMDP**
Safe	Linear-time	EXPTIME-c	EXPTIME-c	Linear-time	EXPTIME-c	EXPTIME-c
Reach	Linear-time	EXPTIME-c	EXPTIME-c	Linear-time	EXPTIME-c	Open
Büchi	Quad-time	EXPTIME-c	EXPTIME-c	Quad-time	EXPTIME-c	Open
co-Büchi	Quad-time	EXPTIME-c	EXPTIME-c	Quad-time	Undecidable	Undecidable
ω-regular	NP $\cap$ coNP	EXPTIME-c	EXPTIME-c	NP $\cap$ coNP	Undecidable	Undecidable
	Limit Sure Winning (Table 4)			**Quantitative Winning**(Table 5)		
	RMDP	**1s RPOMDP**	**RPOMDP**	**RMDP**	**1s RPOMDP**	**RPOMDP**
Safe	Linear-time	EXPTIME-c	EXPTIME-c	NP $\cap$ coNP	Undecidable	Undecidable
Reach	Linear-time	Undecidable	Undecidable	NP $\cap$ coNP	Undecidable	Undecidable
Büchi	Quad-time	Undecidable	Undecidable	NP $\cap$ coNP	Undecidable	Undecidable
co-Büchi	Quad-time	Undecidable	Undecidable	NP $\cap$ coNP	Undecidable	Undecidable
ω-regular	NP $\cap$ coNP	Undecidable	Undecidable	NP $\cap$ coNP	Undecidable	Undecidable

Table 1: Summary of complexity results that follow from our equivalence of RPOMDPs and POSGs under ω-regular objectives. We show results both for general ω-regular objectives as well as for special instances of objectives (reachability, safety, Büchi, co-Büchi) for which more efficient algorithms are known. The '-c' suffix denotes completeness, otherwise shown results denote upper complexity bounds. The red colored entries are our novel results that were previously unknown and that follow from our equivalence result with POSGs [7]. The table considers RMDPs (full observability for both the agent and the environment), 1s-RPOMDPs (partial observability for the agent only), and RPOMDPs (partial observability for both). The blue colored entries remain open problems since the decidability and complexity of solving POSGs under almost sure winning reachability and Büchi objectives is a well-known open problem [7]. The table numbers correspond to the tables in [7] from which the respective complexity results on POSGs were obtained.

the structure of uncertainty sets such as interval MDPs [9, 34], or is restricted to probability 1 or limit-sure satisfaction of ω-regular specifications [2] where the two objectives are shown to have the same value, but without exact complexity bounds that we establish. The one-sided POSG notion we consider is from [7] where observations are on states; [21] considers a notion of one sided stochastic games where observations are on transitions.

The existence of a reduction from R(PO)MDPs to (PO)SGs under reward objectives has been observed in prior work [30, 18]. The work [8] formalizes the reduction from RMDPs to turn-based stochastic games under both expected discounted-sum and long-run average objectives. The work [5] formalizes the reduction from RPOMDPs to POSGs under discounted-sum objectives. However, none of these works consider ω-regular objectives. Moreover, we for the first time establish a polynomial-time reduction in the opposite direction – *from (PO)SGs to R(PO)MDPs* – under ω-regular objectives.

2 Model Definition and Preliminaries

For a set X, we denote by $\Delta(X)$ the set of all probability distributions over X and by 2^X the set of all subsets of X. For $x \in X$, $\mathrm{Dirac}(x) \in \Delta(X)$ is the Dirac distribution assigning probability 1 to x.

For a sequence $b = b_0, b_1, b_2, \cdots, b_n$, we write $\mathrm{last}(b) = b_n$. For $x = (x_1, x_2, \cdots x_k)$, $x(i)$ denotes the element at i-th index. Let $\mathbb{N}$ represent the set of natural numbers.

RPOMDPs. A *Robust POMDP (RPOMDP)* is a tuple $(S, A, \mathcal{P}, \mathcal{O}_\mathsf{a}, \mathcal{O}_\mathsf{e}, \mathcal{Z}_\mathsf{a}, \mathcal{Z}_\mathsf{e}, s_0)$, where S and A are finite sets of states and actions, $s_0 \in S$ is the initial state, and $\mathcal{P} \subseteq (S \times A \to \Delta(S))$, called the *uncertainty set*, is a subset of all functions from $S \times A$ to $\Delta(S)$. $\mathcal{O}_\mathsf{a}, \mathcal{O}_\mathsf{e}$ respectively are finite sets of *observations* for the agent and environment, $\mathcal{Z}_\mathsf{a} : S \to \mathcal{O}_\mathsf{a}, \mathcal{Z}_\mathsf{e} : S \to \mathcal{O}_\mathsf{e}$ are *observation functions* mapping states to observations. A *robust MDP (RMDP)* arises when both the agent and the environment have full observability: for $\star \in \{a, e\}$, $\mathcal{O}_\star = \{o^\star_s \mid s \in S\}$ with $\mathcal{Z}_\mathsf{a}(s) = o^a_s$ and $\mathcal{Z}_\mathsf{e}(s) = o^e_s$. A *one-sided RPOMDP* has partial observability for the agent but full for the environment, i.e. $\mathcal{O}_\mathsf{e} = \{o_s \mid s \in S\}$ with $\mathcal{Z}_\mathsf{e}(s) = o_s$.

Assumptions: (s, a)-rectangularity and Polytopic Uncertainty Sets. As is common in the algorithmic study of RMDPs and uncertain POMDPs [30, 22, 4], we restrict our attention to RPOMDPs with *(s, a)-rectangular uncertainty sets*. That is, we assume that the uncertainty set is of the form $\mathcal{P} = \prod_{(s,a) \in S \times A} P(s, a)$ with each $P(s, a) \subseteq \Delta(S)$, meaning that the environment can choose tran-

sition probabilities independently across state-action pairs. Moreover, we assume that each $P(s,a)$ is a *polytope* with a finite vertex set $V_{s,a}$. We take every $V_{s,a}$ to be given explicitly as part of the input. Accordingly, all model sizes and linear-time claims are measured with respect to these explicit vertex lists. A point in $P(s,a)$ is a distribution over S, see Figure 1(a). Standard examples include L^1 and interval uncertainty sets [20, 15]. In the rest of the paper, RPOMDPs refer to (s,a)-rectangular polytopic RPOMDPs.

Semantics of RPOMDPs. Informally, the semantics of RPOMDPs are defined as follows. The process begins in state s_0 with observations $(\mathcal{Z}_\mathsf{a}(s_0), \mathcal{Z}_\mathsf{e}(s_0))$. At each step t, an agent chooses an action a_t; the agent does not know the exact state s_t, but only the agent observation $o^a_t = \mathcal{Z}_\mathsf{a}(s_t)$. In response, using its observation history and a_t, the adversarial environment selects a complete transition assignment $P_t \in \mathcal{P}$; the transition from the actual state uses its component $P_t(s_t,a_t) \in P(s_t,a_t)$. The environment observes only $o^e_t = \mathcal{Z}_\mathsf{e}(s_t)$. The next state s_{t+1} is sampled from $P_t(s_t,a_t)$, the agent observes $\mathcal{Z}_\mathsf{a}(s_{t+1})$ and the process continues. This process is formalized via agent and environment strategies. Following the terminology of [5], our semantics model *dynamic uncertainty* where the environment may select a new probability assignment at each time step. A *history* is a finite sequence $h_t = s_0, (o^a_0, o^e_0), s_1, \ldots, s_t, (o^a_t, o^e_t) \in (S \times \mathcal{O}_\mathsf{a} \times \mathcal{O}_\mathsf{e})^*$. An *agent (resp., environment) observation history* for h_t is a sequence $oh^a_t = o^a_0, \ldots, o^a_t \in (\mathcal{O}_\mathsf{a})^* \times \mathcal{O}_\mathsf{a}$(resp., $oh^e_t = o^e_0, \ldots, o^e_t \in (\mathcal{O}_\mathsf{e})^* \times \mathcal{O}_\mathsf{e}$). Let OH_a and OH_e denote the sets of agent and environment observation histories. An *agent* and *environment strategy* are mappings

$$\sigma : \mathsf{OH}_a \to \bigcup_{o \in \mathcal{O}_\mathsf{a}} \prod_{s:\, \mathcal{Z}_\mathsf{a}(s)=o} \Delta(A), \qquad \pi : \mathsf{OH}_e \times A \to \mathcal{P},$$

such that $\sigma(oh) \in \prod_{s:\, \mathcal{Z}_\mathsf{a}(s)=\mathrm{last}(oh)} \Delta(A)$. Given a RPOMDP $\mathcal{M}$, let $\Sigma_\mathcal{M}$ and $\Pi_\mathcal{M}$ denote all agent and environment strategies. Note that our strategies are "action invisible" since they map observation histories which do not record actions. A strategy is *finite-memory* if it can be implemented by a finite-state transducer that reads the player's observation history and outputs the required assignment. It is *memoryless* if one memory state suffices; strategies not restricted to finite memory are called *infinite-memory* strategies. A pair of strategies $(\sigma,\pi) \in \Sigma_\mathcal{M} \times \Pi_\mathcal{M}$ induces a *play* $\rho = s_0, a_0, s_1, a_1, \cdots \in (S \times A)^\omega$ starting from s_0. At time t, let $\alpha_t = \sigma(oh^a_t)$ and $P_t = \pi(oh^e_t, a_t)$ be the assignments selected from the players' observation histories. The action a_t is sampled according to the distribution $\alpha_t(s_t)$. The next state satisfies $\Pr(s_{t+1} \mid h_t, a_t) = P_t(s_t,a_t)(s_{t+1})$. Let $\mathsf{Plays}_\mathcal{M}$ denote the set of all plays in $\mathcal{M}$. Given an initial state s_0, each pair of strategies $(\sigma,\pi) \in \Sigma_\mathcal{M} \times \Pi_\mathcal{M}$ induces a probability measure $\mathbb{P}^{\sigma,\pi}_{s_0}(\cdot)$ over the set $\mathsf{Plays}_\mathcal{M}$ [3]. We write $\mathsf{Plays}^{\sigma,\pi}_\mathcal{M}$ for the set of *supported plays*: those plays for which every finite prefix has positive probability under (σ,π).

ω-regular Objectives. A logical objective in an RPOMDP $\mathcal{M}$ is a measurable set $W \subseteq \mathsf{Plays}_\mathcal{M}$. We consider ω-regular objectives [10], specified via a parity automaton. Standard examples are reachability (visit a target state at least once), safety (avoid a bad state forever), Büchi (visit a target state infinitely often), and co-Büchi (visit a bad state only finitely often). It is a classical result in formal methods that every ω-regular objective can be represented via a parity automaton [41]. We assume we have already taken the product of the RPOMDP and the parity automaton, and that the RPOMDP is equipped with a priority function $p : S \to \{0,1,\ldots,d\}, d \in \mathbb{N}$ assigning a priority to each state. The ω-regular objective is then the set of plays in which the minimum priority that is visited infinitely often is even, i.e. $\mathrm{Parity}_\mathcal{M}(p) = \{\rho \in \mathsf{Plays}_\mathcal{M} | \min\{p(s) : s \in \mathrm{Inf}(\rho)\} \text{ is even}\}$. Here $\mathrm{Inf}(\rho)$ is the set of states that occur infinitely often in ρ.

Partially Observable Stochastic Games (POSGs). A turn-based, two-player (max, min) *POSG* [7] is a tuple $\mathcal{G} = (S_\mathsf{max}, S_\mathsf{min}, A_\mathsf{max}, A_\mathsf{min}, \delta, s_0, \mathcal{O}_\mathsf{max}, \mathcal{O}_\mathsf{min}, \mathcal{Z}_\mathsf{max}, \mathcal{Z}_\mathsf{min})$, where S_max and S_min are the sets of max- and min-states, respectively. Write $S^\mathcal{G} = S_\mathsf{max} \cup S_\mathsf{min}$, with initial state $s_0 \in S^\mathcal{G}$. The action sets of the two players are A_max and A_min, and we write $A^\mathcal{G} = A_\mathsf{max} \cup A_\mathsf{min}$. The transition function is $\delta : (S_\mathsf{max} \times A_\mathsf{max}) \cup (S_\mathsf{min} \times A_\mathsf{min}) \to \Delta(S^\mathcal{G})$. Observations are given by sets $\mathcal{O}_\mathsf{max}, \mathcal{O}_\mathsf{min}$ and functions $\mathcal{Z}_\mathsf{max} : S^\mathcal{G} \to \mathcal{O}_\mathsf{max}$, $\mathcal{Z}_\mathsf{min} : S^\mathcal{G} \to \mathcal{O}_\mathsf{min}$ for two players. A POSG is *alternating control (ac-POSG)* if $\delta(s,a) \in \Delta(S_\mathsf{min})$ for every $(s,a) \in S_\mathsf{max} \times A_\mathsf{max}$ and $\delta(s,a) \in \Delta(S_\mathsf{max})$ for every $(s,a) \in S_\mathsf{min} \times A_\mathsf{min}$. Throughout, ac-POSGs start in a max-state, $s_0 \in S_\mathsf{max}$; otherwise, a fresh deterministic initial max-state can be prepended.

A *history* of $\mathcal{G}$ is a sequence $h_t = s_0, (o^1_0, o^2_0), s_1, \ldots, s_t, (o^1_t, o^2_t)$ in $(S^\mathcal{G} \times \mathcal{O}_\mathsf{max} \times \mathcal{O}_\mathsf{min})^*$ such that for all $0 \le i \le t$, $o^1_i = \mathcal{Z}_\mathsf{max}(s_i)$, $o^2_i = \mathcal{Z}_\mathsf{min}(s_i)$. Moreover, for every $0 \le i < t$, there is an action $a \in A_\mathsf{max}$ if $s_i \in S_\mathsf{max}$, or $a \in A_\mathsf{min}$ if $s_i \in S_\mathsf{min}$, such that $\delta(s_i,a)(s_{i+1}) > 0$. For $\star \in \{\mathsf{max}, \mathsf{min}\}$,

if h_t ends in an $S_\star$-state, its $\star$-*observation history* is the sequence of that player's observations $oh^\star = o_0^\star, \ldots, o_t^\star$. We write $\mathsf{OH}_\star^{\mathcal{G}}$ for the set of all such histories. A $\star$-strategy, for $\star \in \{\mathsf{max}, \mathsf{min}\}$, maps each observation history as follows:

$$\sigma_\star : \mathsf{OH}_\star^{\mathcal{G}} \to \bigcup_{o \in \mathcal{O}_\star} \prod_{s \in S_\star :\, \mathcal{Z}_\star(s) = o} \Delta(A_\star),$$

such that $\sigma_\star(oh) \in \prod_{s \in S_\star :\, \mathcal{Z}_\star(s) = \mathrm{last}(oh)} \Delta(A_\star)$. Thus the strategy sees only oh; the game subsequently evaluates the component of the selected assignment belonging to its current state. We write σ and π for the max- and min-strategies, respectively. Note that these are referred to as *action-invisible* strategies in the literature [7]. Plays induced by (σ, π) are defined analogously to RPOMDP plays. $\mathsf{Plays}_{\mathcal{G}}$ denotes the set of all plays. Probability measures of plays, objectives $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ as well as the value $\mathrm{Val}_{\mathcal{G}}(W)$ are defined in a similar manner to RPOMDPs. Let $\Sigma_{\mathcal{G}}$ and $\Pi_{\mathcal{G}}$ denote the sets of all max- and min-strategies, respectively.

Problem. Given an RPOMDP $\mathcal{M}$ with an ω-regular objective W, its *value* and (its value under a fixed agent strategy σ) are defined as $\mathrm{Val}_{\mathcal{M}}(W) = \sup_{\sigma \in \Sigma_{\mathcal{M}}} \inf_{\pi \in \Pi_{\mathcal{M}}} \mathbb{P}_{s_0}^{\sigma,\pi}(W)$ and $(\mathrm{Val}_{\mathcal{M}}^{\sigma}(W) = \inf_{\pi \in \Pi_{\mathcal{M}}} \mathbb{P}_{s_0}^{\sigma,\pi}(W))$, respectively. We consider the four classical ω-regular analysis problems:

- *Almost-sure* analysis asks whether there exists an agent strategy σ such that $\mathrm{Val}_{\mathcal{M}}^{\sigma}(W) = 1$.
- *Limit-sure* analysis asks whether, for every $\epsilon > 0$, there exists an agent strategy σ_ϵ such that $\mathrm{Val}_{\mathcal{M}}^{\sigma_\epsilon}(W) \geq 1 - \epsilon$, equivalently, whether $\mathrm{Val}_{\mathcal{M}}(W) = 1$.
- *Sure* analysis asks whether there exists an agent strategy σ such that $\mathsf{Plays}_{\mathcal{M}}^{\sigma,\pi} \subseteq W$ for every environment strategy π.
- *Quantitative* analysis asks, for a given probability threshold $p \in [0, 1]$, if $\mathrm{Val}_{\mathcal{M}}(W) \geq p$.

The computational analysis problems above have been investigated for POSGs [7].

3 Equivalence with Partially Observable Stochastic Games (ac-POSG)

This section presents the main result of the paper. We provide reductions between RPOMDPs and alternating control POSGs for each problem above. Throughout, we make the standard assumption that every action in a player's action alphabet is enabled at every state owned by that player. An action is *effective* at a state if the construction explicitly specifies its transition there, and an *effective edge* is a positive-probability transition under an effective action. We only display the effective actions and edges; every omitted state-action pair is implicitly completed by an absorbing outcome losing for its owner. This completion does not affect the analysis problems below.

Both reductions are linear in the model size (measured in the number of states, effective actions and transitions).

3.1 Reduction from RPOMDPs to ac-POSGs

Our reduction from RPOMDPs to ac-POSGs draws insight from and generalizes the reduction from RMDPs to stochastic games that was presented in [8]. In this work, we generalize this reduction to the setting with partial observability. In what follows, given a polytopic RPOMDP $\mathcal{M} = (S, A, \mathcal{P}, \mathcal{O}_{\mathsf{a}}, \mathcal{O}_{\mathsf{e}}, \mathcal{Z}_{\mathsf{a}}, \mathcal{Z}_{\mathsf{e}}, s_0)$ with any of the objectives defined in Section 2, we outline our construction of the equivalent ac-POSG $\mathcal{G} = (S_{\mathsf{max}}^{\mathcal{G}}, S_{\mathsf{min}}^{\mathcal{G}}, A_{\mathsf{max}}^{\mathcal{G}}, A_{\mathsf{min}}^{\mathcal{G}}, \delta^{\mathcal{G}}, s_0^{\mathcal{G}}, \mathcal{O}_{\mathsf{max}}^{\mathcal{G}}, \mathcal{O}_{\mathsf{min}}^{\mathcal{G}}, \mathcal{Z}_{\mathsf{max}}^{\mathcal{G}}, \mathcal{Z}_{\mathsf{min}}^{\mathcal{G}})$ and the corresponding objective. The formal construction is deferred to Appendix A.

POSG Construction. The POSG $\mathcal{G}$ is defined as follows. Intuitively, the ac-POSG models the interaction between the agent and the environment in the original RPOMDP, where the max-player corresponds to the agent, the min-player to the environment, and the two players alternate in moves:

States: max-player states $S_{\mathsf{max}}^{\mathcal{G}} = S$ are the states of the original RPOMDP, min-player states $S_{\mathsf{min}}^{\mathcal{G}}$ are state–action pairs $(s, a) \in S \times A$, and the initial state is $s_0^{\mathcal{G}} = s_0 \in S_{\mathsf{max}}^{\mathcal{G}}$. *Actions:* the max-player's actions are those of the RPOMDP, $A_{\mathsf{max}}^{\mathcal{G}} = A$. At min-state (s, a) the effective actions are the vertices $V_{s,a}$, and the global min-action alphabet is $A_{\mathsf{min}}^{\mathcal{G}} = \bigcup_{(s,a) \in S \times A} V_{s,a}$. *Observations:* $\mathcal{O}_{\mathsf{max}}^{\mathcal{G}} = \mathcal{O}_{\mathsf{a}}$ and $\mathcal{O}_{\mathsf{min}}^{\mathcal{G}} = \mathcal{O}_{\mathsf{e}}$, with $\mathcal{Z}_{\mathsf{max}}^{\mathcal{G}}(s) = \mathcal{Z}_{\mathsf{max}}^{\mathcal{G}}((s, a)) = \mathcal{Z}_{\mathsf{a}}(s)$ and $\mathcal{Z}_{\mathsf{min}}^{\mathcal{G}}(s) = \mathcal{Z}_{\mathsf{min}}^{\mathcal{G}}((s, a)) = \mathcal{Z}_{\mathsf{e}}(s)$. *Transitions:* from max-state s under action a, the POSG moves deterministically to the min-state (s, a), i.e. $\delta^{\mathcal{G}}(s, a) = \mathrm{Dirac}((s, a))$; from min-state $s' = (s, a)$ under an effective action $v \in V_{s,a}$,

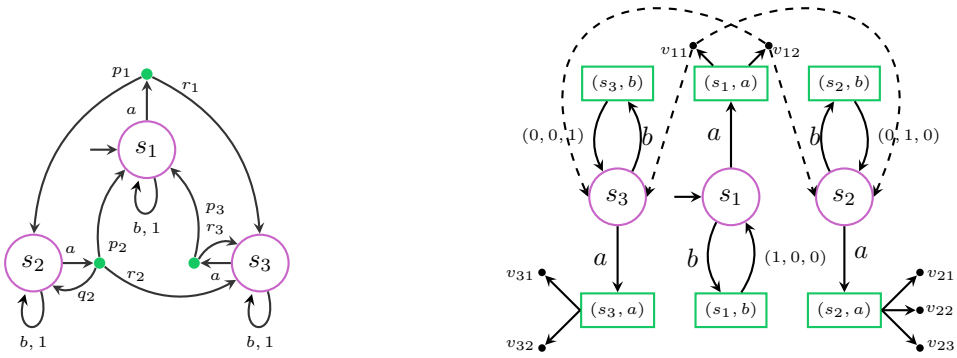

Figure 1: (a) RPOMDP $\mathcal{M}$ with states s_1, s_2, s_3 and actions a, b. Each transition arrow consists of two parts, the first denotes an action (a or b) whereas the second denotes the probability of moving to a state upon taking that action (e.g. p_1 is the probability of moving to s_2 upon taking action a in s_1). The uncertainty set of the RPOMDP is defined by specifying the vertex set of the uncertainty polytope for each state-action pair. We let $V_{s_1,a}$=$\{v_{11}, v_{12}\}$ and $(0, p_1, r_1) \in P(s_1, a)$. $V_{s_2,a}$=$\{v_{21}$,v_{22},$v_{23}\}$ and $(p_2, q_2, r_2) \in P(s_2, a)$. $V_{s_3,a}$=$\{v_{31}, v_{32}\}$ and $(p_3, 0, r_3) \in P(s_3, a)$. Note that on b, all transitions are deterministic, so $V_{s_1,b} = \{(1, 0, 0)\}$, $V_{s_2,b} = \{(0, 1, 0)\}$ and $V_{s_3,b} = \{(0, 0, 1)\}$. (b) The ac-POSG constructed from the RPOMDP, circular and rectangular states represent max-player and min-player states, respectively. The action set of the max-player is the same as the action set in the RPOMDP on the left, namely $\{a, b\}$. Min-player states are all combinations of state-action pairs in the RPOMDP on the left. The effective actions at each such state are the vertices of the corresponding uncertainty polytope. For instance, at (s_1, a) these are v_{11} and v_{12} from $V_{s_1,a}$; randomizing with probabilities (α_1, α_2) mimics the point $\alpha_1 v_{11} + \alpha_2 v_{12}$ in $P(s_1, a)$. We omit the remaining completed actions and the effective edges from (s_2, a) and (s_3, a) for clarity.

$\delta^{\mathcal{G}}(s', v) = v$ is the distribution over $S^{\mathcal{G}}_{\mathsf{max}}$ specified by the vertex v of the uncertainty polytope $P(s, a)$. Randomizing over effective min-actions therefore produces every distribution in $P(s, a)$.

The constructed POSG is *linear* in the size of the RPOMDP and the number of vertices of the uncertainty polytopes. Figure 1(b) shows an example of our construction.

Parity Objectives: The ω-regular objective in the ac-POSG $\mathcal{G}$ is specified by the priority function $p' : S^{\mathcal{G}} \to \{0, 1, \ldots, d\}$ defined via $p'(s') = p(s)$ for any $s' \in S^{\mathcal{G}}$, where s is the state component of s' ($s' = s$ or $s' = (s, a)$) and p is the priority function of $\mathcal{M}$.

Theorem 3.1 (Proof in Appendix A.4)**.** *Let $\mathcal{M}$ be a RPOMDP and $\mathcal{G}$ the induced ac-POSG. Then $\mathrm{Val}_{\mathcal{M}}(W) = \mathrm{Val}_{\mathcal{G}}(W')$ for any ω-regular objective W in $\mathcal{M}$ and corresponding objective W' in $\mathcal{G}$. Consequently, sure, almost-sure, limit-sure, and quantitative analysis of any objective above on $\mathcal{M}$ is inter-reducible in linear time, and therefore in polynomial time, with the corresponding analysis on $\mathcal{G}$, under the explicit vertex representation fixed in Section 2.*

3.2 Reduction from ac-POSGs to RPOMDPs

Key Challenge and Our Solution. Reduction from ac-POSGs to RPOMDPs turns out to be much more challenging due to the following subtle reason. Given an ac-POSG, if the max-player in a state s chooses action a, the min-player gets to see the observation $\mathcal{Z}_{\mathsf{min}}(s')$ where $\delta(s, a)(s') > 0$ before choosing its own action. On the contrary, in an RPOMDP, after the agent chooses an action a at a state s, the environment at (s, a) commits to a polytope $P(s, a)$ immediately, with nothing observed in between. Thus, at a first glance, it seems that the min-player in ac-POSGs has more power compared to the environment in RPOMDPs. Somewhat surprisingly, we show that this is not the case and that it is still possible to construct a reduction from ac-POSGs to RPOMDPs. This observation motivates us to construct a two-step reduction. The first step of the reduction reduces ac-POSGs to an intermediate model which we call *pre-min-transformed POSG*, which resolves this semantic discrepancy. The second step then reduces a pre-min-transformed POSG to an RPOMDP.

3.2.1 Step 1: Reduction from ac-POSGs to Pre-min-transformed POSGs

The pre-min transformation inserts, before every min-state $s \in S_{\mathsf{min}}$, a fresh *dummy* max-state d_s whose sole effective action leads deterministically to s, and whose observations to both players are copies of the observations of s. Neither player gains strategic power: the max-player has nothing to choose at d_s (every other globally enabled action takes the max-losing default transition), while the min-player sees the same information at s as in $\mathcal{G}$, the observation at d_s is determined by the observation at s and so reveals nothing new. The following construction formalizes this intuition.

Pre-min-transformed POSG. Let $\mathcal{G} = (S_{\mathsf{max}}, S_{\mathsf{min}}, A_{\mathsf{max}}, A_{\mathsf{min}}, \delta, s_0, \mathcal{O}_{\mathsf{max}}, \mathcal{O}_{\mathsf{min}}, \mathcal{Z}_{\mathsf{max}}, \mathcal{Z}_{\mathsf{min}})$ be an ac-POSG, $D = \{d_s \mid s \in S_{\mathsf{min}}\}$ be the set of dummy states , and $\mathcal{O}^\dagger = \{o^\dagger_x \mid x \in \mathcal{O}_{\mathsf{max}}\}$ and $\mathcal{O}^\# = \{o^\#_x \mid x \in \mathcal{O}_{\mathsf{min}}\}$ be copies of observations for both players. The corresponding pre-min-transformed POSG $\mathcal{G}' = (S'_{\mathsf{max}}, S'_{\mathsf{min}}, A'_{\mathsf{max}}, A'_{\mathsf{min}}, \delta', s_0, \mathcal{O}'_{\mathsf{max}}, \mathcal{O}'_{\mathsf{min}}, \mathcal{Z}'_{\mathsf{max}}, \mathcal{Z}'_{\mathsf{min}})$ is defined via:

- *States.* $S'_{\mathsf{max}} = S_{\mathsf{max}} \cup D$ and $S'_{\mathsf{min}} = S_{\mathsf{min}}$, with initial state s_0.
- *Actions.* $A'_{\mathsf{max}} = A_{\mathsf{max}} \cup \{\top\}$ and $A'_{\mathsf{min}} = A_{\mathsf{min}}$, with all respective actions enabled throughout. The effective max-actions are A_{max} at original max-states and the fresh action $\top$ at dummy states.
- *Observations.* $\mathcal{O}'_{\mathsf{max}} = \mathcal{O}_{\mathsf{max}} \cup \mathcal{O}^\dagger$ and $\mathcal{O}'_{\mathsf{min}} = \mathcal{O}_{\mathsf{min}} \cup \mathcal{O}^\#$, with $\mathcal{Z}'_{\mathsf{max}}(s) = \mathcal{Z}_{\mathsf{max}}(s)$ and $\mathcal{Z}'_{\mathsf{min}}(s) = \mathcal{Z}_{\mathsf{min}}(s)$ for all $s \in S_{\mathsf{max}} \cup S_{\mathsf{min}}$, while $\mathcal{Z}'_{\mathsf{max}}(d_s) = o^\dagger_{\mathcal{Z}_{\mathsf{max}}(s)}$ and $\mathcal{Z}'_{\mathsf{min}}(d_s) = o^\#_{\mathcal{Z}_{\mathsf{min}}(s)}$ for all $d_s \in D$ are the newly introduced observation copies.
- *Transition function.* (i) For $s_1 \in S_{\mathsf{max}}$, $a_1 \in A_{\mathsf{max}}$, and $s_2 \in S_{\mathsf{min}}$ such that $\delta(s_1, a_1)(s_2) > 0$, the pre-min-transformed POSG transitions to the min-player copy d_{s_2} rather than to the min-player state, i.e. $\delta'(s_1, a_1)(d_{s_2}) = \delta(s_1, a_1)(s_2)$. (ii) For $s_2 \in S_{\mathsf{min}}$ and $a_2 \in A_{\mathsf{min}}$, the transition function is the same as in the original ac-POSG, i.e. $\delta'(s_2, a_2) = \delta(s_2, a_2)$. (iii) For the newly introduced copy states $d_s \in D$ and action $\top$, we define $\delta'(d_s, \top) = \mathrm{Dirac}(s)$.

Pre-min-transformed POSG objective. Plays of $\mathcal{G}$ and $\mathcal{G}'$ are related by a bijection $\Gamma : \mathsf{Plays}_{\mathcal{G}} \to \mathsf{Plays}_{\mathcal{G}'}$. Since $\mathcal{G}$ is alternating control, every $h = s_0, a_0, s_1, a_1, \cdots \in \mathsf{Plays}_{\mathcal{G}}$ has $s_{2k} \in S_{\mathsf{max}}$ and $s_{2k+1} \in S_{\mathsf{min}}$, and Γ inserts the pair $(d_{s_{2k+1}}, \top)$ between every S_{max}-step and the following S_{min}-step. Observation histories lift the same way via $\Gamma^{\mathsf{max}}_{\mathcal{O}}, \Gamma^{\mathsf{min}}_{\mathcal{O}}$. Strategies in $\mathcal{G}, \mathcal{G}'$ are functions of observation histories. We lift them between the two games as follows. For $\star \in \{\mathsf{max}, \mathsf{min}\}$, $\Phi_\star$ takes a $\mathcal{G}$-strategy σ to the $\mathcal{G}'$-strategy that, given a $\mathcal{G}'$-observation history oh' ending in non-dummy state, strips the inserted observations and queries σ at the underlying $\mathcal{G}$-history: $\Phi_\star(\sigma)(oh') = \sigma((\Gamma^\star_O)^{-1}(oh'))$; for ending at dummy states $\Phi_{\mathsf{max}}(\sigma)$ plays $\top$ deterministically. The reverse $\Psi_\star$ goes the other way: $\Psi_\star(\sigma')(oh) = \sigma'(\Gamma^\star_O(oh))$. The pairs are mutual inverses, $\Phi_\star = (\Psi_\star)^{-1}$, which helps us lift objectives across to $\mathcal{G}'$.

• **Parity objective.** Let $p : S_{\mathsf{max}} \cup S_{\mathsf{min}} \to \{0, 1, \ldots, d\}$ be the priority function for $\mathcal{G}$. Define $p' : S'_{\mathsf{max}} \cup S'_{\mathsf{min}} \to \{0, 1, \ldots, d\}$ by $p'(s) = p(s)$ for $s \in S_{\mathsf{max}} \cup S_{\mathsf{min}}$ and $p'(d_s) = p(s)$ for $d_s \in D$.

Size and memory preservation under Pre-min. It is easy to see that $|\mathcal{G}'| = \mathcal{O}(|\mathcal{G}|)$ and the construction runs in linear time. Moreover, $|\Gamma^\star_O(oh)|$ $(|\Gamma^\star_O)^{-1}(oh')|)$ is linear in $|oh|$ $(|oh'|)$. Finally, the lifting maps $\Phi_\star, \Psi_\star$ preserve memory class exactly: memoryless, finite-memory and infinite-memory regimes correspond exactly between $\mathcal{G}$ and $\mathcal{G}'$. See details in Appendix B.3.1.

Lemma 3.2 (Equivalence between $\mathcal{G}$ and $\mathcal{G}'$, Appendix B.4)**.** *Let $\mathcal{G}$ be an ac-POSG and $\mathcal{G}'$ its Pre-*min *transformation. The play map Γ and the strategy maps $\Phi_\star, \Psi_\star$ ($\star \in \{\mathsf{max}, \mathsf{min}\}$) satisfy the following: for every priority function p on $\mathcal{G}$, every ω-regular objective W in $\mathcal{G}$ and the lifted objective W' in $\mathcal{G}'$ and every $\rho \in \mathsf{Plays}_{\mathcal{G}}, \sigma \in \Sigma_{\mathcal{G}}, \pi \in \Pi_{\mathcal{G}}$:*

(I, obj.) $\rho \in W \iff \Gamma(\rho) \in W'$; (II, plays) $\Gamma(\mathsf{Plays}^{\sigma,\pi}_{\mathcal{G}}) = \mathsf{Plays}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}$;

(III, prob.) $\mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W) = \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(W')$; (IV, val.) $\mathrm{Val}_{\mathcal{G}}(W) = \mathrm{Val}_{\mathcal{G}'}(W')$.

Consequently, sure, almost-sure, limit-sure, and quantitative analysis of any objective above on $\mathcal{G}$ is inter-reducible in linear time with the corresponding analysis on $\mathcal{G}'$.

3.2.2 Step 2: Reduction from Pre-min-transformed POSGs to RPOMDPs

We translate a Pre-min-transformed POSG $\mathcal{G}'$ into a value-equivalent RPOMDP $\mathfrak{R}(\mathcal{G})$ by *eliminating the* min*-player's states* and folding the min-player's choices into polytope choices for the RPOMDP environment. In $\mathcal{G}'$, every play has the periodic structure $s_1 \xrightarrow{a_1} d_{s_2} \xrightarrow{\top} s_2 \xrightarrow{a_2} s_3 \xrightarrow{a_3} d_{s_4} \xrightarrow{\top}$

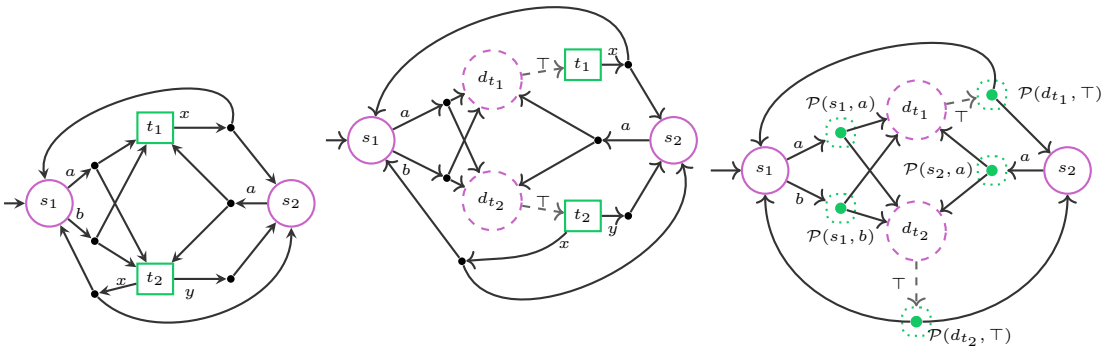

Figure 2: (a) A POSG $\mathcal{G}$. max-states s_1, s_2 are circles, while min-states t_1, t_2 are rectangles. (b) The Pre-min transformed game $\mathcal{G}'$ with dashed circles representing new dummy states- d_{t_1}, d_{t_2} and dashed arrows represent the $\top$-action. (c) The reduced RPOMDP $\mathfrak{R}(\mathcal{G})$ with green dotted circles represent uncertainty sets $P(s, a)$. $P(s_1, a)$ is a singleton consisting of the distribution $\delta(s_1, a)$, likewise for $P(s_1, b)$ and $P(s_2, a)$. $P(d_{t_1}, \top)$ is also a singleton $\{\delta'(d_{t_1}, x)\}$. $P(d_{t_2}, \top)$ is the polytope with two vertices $\{\delta'(d_{t_2}, x), \delta'(d_{t_2}, y)\}$.

$s_4 \cdots$, in which $S_{\mathsf{max}} \cup D$ states alternate with S_{min}-states. The key observation is that the min-player's only role is to pick an action $a \in A_{\mathsf{min}}$ at each S_{min}-state s, and this choice deterministically fixes a transition distribution $\delta'(s, a) \in \Delta(S_{\mathsf{max}})$. This is precisely the role of the environment in an RPOMDP: given a state and an action, it selects a distribution from the uncertainty set. We exploit this correspondence by collapsing each S_{min}-state s into the dummy state d_s that immediately precedes it: the polytope $P(d_s, \top)$ has as its vertices, the distributions $\delta'(s, a)$, one per min-action $a \in A_{\mathsf{min}}$. Picking a point in this polytope is a convex combination $\sum_a \alpha_a \, \delta'(s, a)$, where α_a is the distribution value induced by min$-$action a at s. The RPOMDP environment at $(d_s, \top)$ therefore has exactly the choices available to the min-player at s in $\mathcal{G}'$. The following construction formalizes this intuition.

RPOMDP construction. Given a Pre-min-transformed POSG $\mathcal{G}' = (S_{\mathsf{max}} \cup D, S_{\mathsf{min}}, A_{\mathsf{max}} \cup \{\top\}, A_{\mathsf{min}}, \delta', s_0, \mathcal{O}_{\mathsf{max}} \cup \mathcal{O}^{\dagger}, \mathcal{O}_{\mathsf{min}} \cup \mathcal{O}^{\#}, \mathcal{Z}'_{\mathsf{max}}, \mathcal{Z}'_{\mathsf{min}})$, the corresponding RPOMDP $\mathfrak{R}(\mathcal{G}) = (S_{\mathfrak{R}}, A_{\mathfrak{R}}, \mathcal{P}_{\mathfrak{R}}, \mathcal{O}^{\mathfrak{R}}_{\mathsf{a}}, \mathcal{O}^{\mathfrak{R}}_{\mathsf{e}}, \mathcal{Z}^{\mathfrak{R}}_{\mathsf{a}}, \mathcal{Z}^{\mathfrak{R}}_{\mathsf{e}}, s^{\mathfrak{R}}_0)$ is defined as follows:

- *States.* $S_{\mathfrak{R}} = S_{\mathsf{max}} \cup D$, with initial state $s^{\mathfrak{R}}_0 = s_0$.
- *Actions.* $A_{\mathfrak{R}} = A_{\mathsf{max}} \cup \{\top\}$, with every action enabled at every RPOMDP state. The construction lists A_{max} as effective at original max-states and $\top$ at dummy states.
- *Uncertainty set.* $\mathcal{P}_{\mathfrak{R}} = \prod_{(s,a) \in S_{\mathfrak{R}} \times A_{\mathfrak{R}}} P(s, a)$. The polytopes have vertex sets $V_{s,a} = \{\delta'(s, a)\}$ if $s, a \in S_{\mathsf{max}} \times A_{\mathsf{max}}$, $V_{s,a} = \{\delta'(s', a') \mid a' \in A_{\mathsf{min}}\}$ if $s = d_{s'}$, $a = \top$.
- *Observations.* $\mathcal{O}^{\mathfrak{R}}_{\mathsf{a}} = \mathcal{O}_{\mathsf{max}} \cup \mathcal{O}^{\dagger}$ and $\mathcal{O}^{\mathfrak{R}}_{\mathsf{e}} = \mathcal{O}_{\mathsf{min}} \cup \mathcal{O}^{\#}$, with $\mathcal{Z}^{\mathfrak{R}}_{\mathsf{a}}(s') = \mathcal{Z}'_{\mathsf{max}}(s')$ and $\mathcal{Z}^{\mathfrak{R}}_{\mathsf{e}}(s') = \mathcal{Z}'_{\mathsf{min}}(s')$ for all $s' \in S_{\mathfrak{R}}$.

RPOMDP objectives. We first relate plays of $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$. For $\rho' \in \mathsf{Plays}_{\mathcal{G}'}$, $\Lambda : \mathsf{Plays}_{\mathcal{G}'} \to \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}$ removes every S_{min}-state and every A_{min}-action from ρ'. The map Λ is *many-to-one*: $\mathcal{G}'$-plays that agree on $S_{\mathsf{max}} \cup D$ states and $A_{\mathsf{max}} \cup \{\top\}$ actions but differ in the A_{min}-action chosen at S_{min}-states map to the same $\mathfrak{R}(\mathcal{G})$-play. Its set-valued inverse $\Lambda^{-1}(\rho^{\mathfrak{R}})$ recovers the pre-image by re-introducing an arbitrary A_{min}-action consistent with the realised next state. Observation histories lift via $\Lambda^{\mathsf{max}}_{\mathcal{O}}, \Lambda^{\mathsf{min}}_{\mathcal{O}}$: the agent map $\Lambda^{\mathsf{max}}_{\mathcal{O}}$ is a bijection between $\mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$ and $\mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}}$, while the environment map $\Lambda^{\mathsf{min}}_{\mathcal{O}}$ is a bijection onto the co-domain of environment histories ending at a dummy state.

Strategies in $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$ are functions of observation histories. We lift strategies between the two models as follows. $\Omega_{\mathsf{max}} : \Sigma_{\mathcal{G}'} \to \Sigma_{\mathfrak{R}(\mathcal{G})}$ converts a $\mathcal{G}'$-max-strategy σ' into an agent strategy. Given an agent observation history $oh^{\mathfrak{R}}$, the new strategy first lifts $oh^{\mathfrak{R}}$ to its $\mathcal{G}'$-counterpart $(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}})$ and queries σ' there: $\Omega_{\mathsf{max}}(\sigma')(oh^{\mathfrak{R}}) = \sigma'\big((\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}})\big)$. $\Omega_{\mathsf{min}} : \Pi_{\mathcal{G}'} \to \Pi_{\mathfrak{R}(\mathcal{G})}$ converts a $\mathcal{G}'$-min-strategy π' into an environment strategy. At a dummy state $d_{s'}$ under $\top$, with environment history $oh^{\mathfrak{R}}$ ending in $\mathcal{O}^{\#}$, write $\pi'\big((\Lambda^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}})\big)(s')(a) = \alpha_a$ for the randomised min-action a at s'; then $\Omega_{\mathsf{min}}(\pi')$ selects the convex combination $\sum_{a \in A_{\mathsf{min}}} \alpha_a \cdot \delta'(s', a) \in P(d_{s'}, \top)$. At non-dummy

state-action pairs, the polytope is a singleton, and $\Omega_{\mathsf{min}}(\pi')$ picks its unique element. The reverse maps $\Xi_{\mathsf{max}} : \Sigma_{\mathfrak{R}(\mathcal{G})} \to \Sigma_{\mathcal{G}'}$ and $\Xi_{\mathsf{min}} : \Pi_{\mathfrak{R}(\mathcal{G})} \to \Pi_{\mathcal{G}'}$ are defined analogously: $\Xi_{\mathsf{max}}(\sigma^{\mathfrak{R}})(oh') = \sigma^{\mathfrak{R}}(\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh'))$; the environment map is defined analogously from the convex coefficients of the selected point. The formal strategy maps are given in the Appendix.

• **Parity objective.** Let p' be the priority function for $\mathcal{G}'$. Define the priority function for $\mathfrak{R}(\mathcal{G})$, $p^{\mathfrak{R}} : S_{\mathfrak{R}} \to \{0, 1, \ldots, d\}$ by $p^{\mathfrak{R}}(s') = p'(s')$ for all $s' \in S_{\mathfrak{R}}$.

Size and memory preservation under RPOMDP construction. It is easy to see that $|\mathfrak{R}(\mathcal{G})| = \mathcal{O}(|\mathcal{G}'|)$ and the construction runs in linear time. Moreover, observation histories on the two sides are linear in each other. The lifting maps $\Omega_\star, \Xi_\star$ preserve memory class exactly: memoryless, finite-memory, and infinite-memory regimes correspond exactly between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$ (Appendix C.3.1).

Lemma 3.3 (Equivalence between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$, Proof in Appendix C.4)**.** *Let $\mathcal{G}'$ be a Pre-min-transformed POSG and $\mathfrak{R}(\mathcal{G})$ its RPOMDP reduction. The play map Λ and the strategy maps $\Omega_\star, \Xi_\star$ ($\star \in \{\mathsf{max}, \mathsf{min}\}$) satisfy the following: for every priority function p' on $\mathcal{G}'$, every ω-regular objective W' in $\mathcal{G}'$ with the lifted objective $W^{\mathfrak{R}}$ in $\mathfrak{R}(\mathcal{G})$ and for every $\rho' \in \mathsf{Plays}_{\mathcal{G}'}$, $\sigma' \in \Sigma_{\mathcal{G}'}$ and $\pi' \in \Pi_{\mathcal{G}'}$:*
(I, obj.) $\rho' \in W' \iff \Lambda(\rho') \in W^{\mathfrak{R}}$; (II, plays) $\Lambda(\mathsf{Plays}^{\sigma',\pi'}_{\mathcal{G}'}) = \mathsf{Plays}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}$;
(III, prob.) $\mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(W') = \mathbb{P}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}})$; (IV, val.) $\mathrm{Val}_{\mathcal{G}'}(W') = \mathrm{Val}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}})$
Consequently, sure, almost-sure, limit-sure, and quantitative analysis of any objective defined above on $\mathcal{G}'$ is inter-reducible in linear time with the corresponding analysis on $\mathfrak{R}(\mathcal{G})$.

Proof outline. The four clauses are proved in sequence, each using the previous one. (1) holds because Λ leaves the underlying $S_{\mathsf{max}} \cup D$-state sequence unchanged: target states are visited at exactly the same positions, and $p^{\mathfrak{R}}$ is just the restriction of p' to $S_{\mathfrak{R}}$, so the only states removed from ρ' are S_{min}-states whose priority by construction equals that of the preceding dummy state – which does not change the $\liminf$ used in the parity condition. (2): Take $\rho' \in \mathsf{Plays}^{\sigma',\pi'}_{\mathcal{G}'}$ and check, step by step, that $\Lambda(\rho')$ could have been produced by the strategy pair $(\Omega_{\mathsf{max}}(\sigma'), \Omega_{\mathsf{min}}(\pi'))$ in $\mathfrak{R}(\mathcal{G})$. At each S_{max}-state, the singleton uncertainty set forces the unique transition; at each dummy state, the convex combination $\sum_a \alpha_a \delta'(s, a)$ chosen by $\Omega_{\mathsf{min}}(\pi')$ is exactly the next-state distribution induced by π' at s. Conversely, every play in $\mathsf{Plays}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}$ has a $\mathcal{G}'$-pre-image in Λ^{-1} obtained by drawing, at each dummy step, a min-action according to the coefficients (α_a). (3) is a cylinder-set induction on history length with three cases (S_{max}-extension, dummy-state extension, and $S_{\mathsf{min}} \to S_{\mathsf{max}}$ traversal): in each case the probability of the next step in $\mathfrak{R}(\mathcal{G})$ reduces to the corresponding $\mathcal{G}'$-probability, with the deterministic $\top$-transition contributing 1 and the polytope coefficients matching the min-action distribution. (4) first applies (3) in both mapping directions to show equality of the inner infima for each fixed agent strategy, and then takes the outer supremum; only agent maps need to be bijective. The "consequently" is immediate: quantitative and limit-sure analysis follow from clause (4); almost-sure follows from (3) by taking the infimum over π' on each side and matching witnesses through Ω and Ξ; and sure analysis follows from (2) and (1) since a strategy keeping all $\mathcal{G}'$-plays inside W' corresponds via Ω to a strategy keeping all $\mathfrak{R}(\mathcal{G})$-plays inside $W^{\mathfrak{R}}$, and conversely via Ξ. Thus from the above and from Lemmas 3.3 and 3.2, we conclude:

Theorem 3.4 (Equivalence of $\mathcal{G}$ and $\mathfrak{R}(\mathcal{G})$)**.** *Let $\mathcal{G}$ be an ac-POSG and $\mathfrak{R}(\mathcal{G})$ be the constructed RPOMDP as in Section 3.2. The sure, almost-sure, limit-sure, and quantitative analysis of any ω-regular objective on $\mathcal{G}$ is inter-reducible in linear time with the corresponding analysis on $\mathfrak{R}(\mathcal{G})$.*

Complexity Results. A corollary of the reductions in Sections 3.1 and 3.2 is that, under (s, a)-rectangularity, Polytopic RPOMDPs are equivalent to finite ac-POSGs for any ω-regular objective. Our proofs show this for *action-invisible strategies*, where agent and environment strategies cannot observe the agent's chosen actions in their respective observation histories. Thanks to the bidirectional equivalence, RPOMDPs inherit both upper and lower complexity bounds from POSGs for {sure, almost-sure, limit-sure, quantitative} winning under ω-regular objectives, see [7] for a survey. Table 1 summarizes these results.

4 Conclusion

We studied RPOMDPs under ω-regular objectives, and showed that solving them is equivalent to solving ac-POSGs under ω-regular objectives. We do this by establishing polynomial-time reductions

in both directions. Exploiting this equivalence, we derived several new computational complexity results for both RPOMDPs and RMDPs, and provided a complete picture of the complexity landscape of the problem under action invisible strategies. Our results apply to R(PO)MDPs with (s, a)-rectangular and polytopic uncertainty sets. Interesting future work would be studying computational complexity bounds under more general settings.

References

[1] A. Asadi, K. Chatterjee, E. K. Goharshady, M. Karrabi, A. Montaseri, and C. Pagano. Strongly polynomial time complexity of policy iteration for l_∞ robust mdps. *CoRR*, abs/2601.23229, 2026.

[2] A. Asadi, K. Chatterjee, E. K. Goharshady, M. Karrabi, and A. Shafiee. Qualitative analysis of ω-regular objectives on robust mdps. In *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, pages 36137–36145, 2026.

[3] P. Billingsley. *Probability and measure*. John Wiley & Sons, 2017.

[4] E. M. Bovy. The underlying belief model of uncertain partially observable markov decision processes. Master's thesis, Radboud University, 2023.

[5] E. M. Bovy, M. Suilen, S. Junges, and N. Jansen. Imprecise probabilities meet partial observability: Game semantics for robust pomdps. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea, August 3-9, 2024*, pages 6697–6706, 2024.

[6] M. E. Chamie and H. Mostafa. Robust action selection in partially observable markov decision processes with model uncertainty. In *57th IEEE Conference on Decision and Control, CDC 2018, Miami, FL, USA, December 17-19, 2018*, pages 5586–5591, 2018.

[7] K. Chatterjee, L. Doyen, and T. A. Henzinger. A survey of partial-observation stochastic parity games. *Formal Methods Syst. Des.*, 43(2):268–284, 2013.

[8] K. Chatterjee, E. K. Goharshady, M. Karrabi, P. Novotný, and D. Zikelic. Solving long-run average reward robust mdps via stochastic games. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea, August 3-9, 2024*, pages 6707–6715, 2024.

[9] K. Chatterjee, K. Sen, and T. A. Henzinger. Model-checking omega-regular properties of interval markov chains. In *Foundations of Software Science and Computational Structures, 11th International Conference, FOSSACS 2008, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2008, Budapest, Hungary, March 29 - April 6, 2008. Proceedings*, pages 302–317, 2008.

[10] E. M. Clarke, T. A. Henzinger, H. Veith, and R. Bloem, editors. *Handbook of Model Checking*. Springer, 2018.

[11] M. Cubuktepe, N. Jansen, S. Junges, A. Marandi, M. Suilen, and U. Topcu. Robust finite-state controllers for uncertain pomdps. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 11792–11800, 2021.

[12] M. F. L. Galesloot, R. Andriushchenko, M. Ceska, S. Junges, and N. Jansen. Robust finite-memory policy gradients for hidden-model pomdps. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025, Montreal, Canada, August 16-22, 2025*, pages 8518–8526, 2025.

[13] M. F. L. Galesloot, M. Suilen, T. D. Simão, S. Carr, M. T. J. Spaan, U. Topcu, and N. Jansen. Pessimistic iterative planning with rnns for robust pomdps. In *ECAI 2025 - 28th European Conference on Artificial Intelligence, 25-30 October 2025, Bologna, Italy - Including 14th Conference on Prestigious Applications of Intelligent Systems (PAIS 2025)*, pages 4823–4831, 2025.

[14] M. Ghavamzadeh, M. Petrik, and Y. Chow. Safe policy improvement by minimizing robust baseline regret. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 2298–2306, 2016.

[15] R. Givan, S. M. Leach, and T. L. Dean. Bounded-parameter markov decision processes. *Artif. Intell.*, 122(1-2):71–109, 2000.

[16] V. Goyal and J. Grand-Clément. Robust markov decision processes: Beyond rectangularity. *Math. Oper. Res.*, 48(1):203–226, 2023.

[17] J. Grand-Clément and M. Petrik. Reducing blackwell and average optimality to discounted mdps via the blackwell discount factor. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[18] J. Grand-Clément, M. Petrik, and N. Vieille. Beyond discounted returns: Robust markov decision processes with average and blackwell optimality. *CoRR*, abs/2312.03618, 2023.

[19] E. A. Hansen, D. S. Bernstein, and S. Zilberstein. Dynamic programming for partially observable stochastic games. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA*, pages 709–715, 2004.

[20] C. P. Ho, M. Petrik, and W. Wiesemann. Partial policy iteration for l1-robust markov decision processes. *J. Mach. Learn. Res.*, 22:275:1–275:46, 2021.

[21] K. Horák, B. Bošanskỳ, V. Kovařík, and C. Kiekintveld. Solving zero-sum one-sided partially observable stochastic games. *Artificial Intelligence*, 316:103838, 2023.

[22] G. N. Iyengar. Robust dynamic programming. *Math. Oper. Res.*, 30(2):257–280, 2005.

[23] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artif. Intell.*, 101(1-2):99–134, 1998.

[24] D. L. Kaufman and A. J. Schaefer. Robust modified policy iteration. *INFORMS J. Comput.*, 25(3):396–410, 2013.

[25] H. K. Khalil and J. W. Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

[26] M. Krale, E. M. Bovy, M. F. L. Galesloot, T. D. Simão, and N. Jansen. On evaluating policies for robust POMDPs. In *Eighteenth European Workshop on Reinforcement Learning*, 2025.

[27] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas. Temporal-logic-based reactive mission and motion planning. *IEEE Trans. Robotics*, 25(6):1370–1381, 2009.

[28] T. Meggendorfer, M. Weininger, and P. Wienhöft. Solving robust markov decision processes: Generic, reliable, efficient. In *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA, USA, February 25 - March 4, 2025*, pages 26631–26641, 2025.

[29] H. Nakao, R. Jiang, and S. Shen. Distributionally robust partially observable markov decision process with moment-based ambiguity. *SIAM J. Optim.*, 31(1):461–488, 2021.

[30] A. Nilim and L. E. Ghaoui. Robustness in markov decision problems with uncertain transition matrices. In *Advances in Neural Information Processing Systems 16 [Neural Information Processing Systems, NIPS 2003, December 8-13, 2003, Vancouver and Whistler, British Columbia, Canada]*, pages 839–846, 2003.

[31] T. Osogami. Robust partially observable markov decision process. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 106–115, 2015.

[32] A. Pnueli. The temporal logic of programs. In *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*, pages 46–57, 1977.

[33] A. Roy, H. Xu, and S. Pokutta. Reinforcement learning under model mismatch. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 3043–3052, 2017.

[34] K. Sen, M. Viswanathan, and G. Agha. Model-checking markov chains in the presence of uncertainties. In *Tools and Algorithms for the Construction and Analysis of Systems, 12th International Conference, TACAS 2006 Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2006, Vienna, Austria, March 25 - April 2, 2006, Proceedings*, pages 394–410, 2006.

[35] M. Suilen, T. Badings, E. M. Bovy, D. Parker, and N. Jansen. Robust markov decision processes: A place where AI and formal methods meet. In *Principles of Verification: Cycling the Probabilistic Landscape - Essays Dedicated to Joost-Pieter Katoen on the Occasion of His 60th Birthday, Part III*, pages 126–154, 2024.

[36] M. Suilen, N. Jansen, M. Cubuktepe, and U. Topcu. Robust policy synthesis for uncertain pomdps via convex optimization. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4113–4120, 2020.

[37] S. Summers and J. Lygeros. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Autom.*, 46(12):1951–1961, 2010.

[38] A. Tamar, S. Mannor, and H. Xu. Scaling up robust mdps using function approximation. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 181–189, 2014.

[39] C. Tessler, Y. Efroni, and S. Mannor. Action robust reinforcement learning and applications in continuous control. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pages 6215–6224, 2019.

[40] A. Tewari and P. L. Bartlett. Bounded parameter markov decision processes with average reward criterion. In *Learning Theory, 20th Annual Conference on Learning Theory, COLT 2007, San Diego, CA, USA, June 13-15, 2007, Proceedings*, pages 263–277, 2007.

[41] W. Thomas. Languages, automata, and logic. In *Handbook of Formal Languages, Volume 3: Beyond Words*, pages 389–455. 1997.

[42] Y. Wang, A. Velasquez, G. K. Atia, A. Prater-Bennette, and S. Zou. Robust average-reward markov decision processes. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 15215–15223, 2023.

[43] Y. Wang, A. Velasquez, G. K. Atia, A. Prater-Bennette, and S. Zou. Robust average-reward reinforcement learning. *J. Artif. Intell. Res.*, 80:719–803, 2024.

[44] Y. Wang and S. Zou. Online robust reinforcement learning with model uncertainty. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 7193–7206, 2021.

[45] Y. Wang and S. Zou. Policy gradient method for robust reinforcement learning. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, pages 23484–23526, 2022.

[46] W. Wiesemann, D. Kuhn, and B. Rustem. Robust markov decision processes. *Math. Oper. Res.*, 38(1):153–183, 2013.

Appendix

Table of Contents

As in the main text, all actions are formally enabled at all states. When a construction displays only effective actions, omitted actions are implicitly completed by an absorbing outcome losing for their owner. We suppress these dominated transitions and carry out the mappings on the displayed effective actions. For the max-player, assigning positive probability to an omitted action can only decrease the probability of satisfying the objective, so that probability can be reassigned to an effective action. Dually, an omitted min-action reaches a max-winning outcome and therefore cannot improve the min-player's strategy. Thus both players may be restricted to effective actions without changing the value or the sure, almost-sure, limit-sure, and quantitative winning sets. This argument applies because all objectives considered here are represented as parity objectives and the completion outcomes are absorbing wins or losses; an arbitrary action-sensitive objective would first need to be encoded in the state space.

A Reduction of RPOMDPs to POSGs

Definition: Let $\mathcal{M} = (S, A, \mathcal{P}, \mathcal{O}_{\mathsf{a}}, \mathcal{O}_{\mathsf{e}}, \mathcal{Z}_{\mathsf{a}}, \mathcal{Z}_{\mathsf{e}}, s_0)$ be a polytopic RPOMDP. We define the POSG induced by $\mathcal{M}$ as $\mathcal{G} = (S^{\mathcal{G}}_{\mathsf{max}}, S^{\mathcal{G}}_{\mathsf{min}}, A^{\mathcal{G}}_{\mathsf{max}}, A^{\mathcal{G}}_{\mathsf{min}}, \delta^{\mathcal{G}}, s^{\mathcal{G}}_0, \mathcal{O}^{\mathcal{G}}_{\mathsf{max}}, \mathcal{O}^{\mathcal{G}}_{\mathsf{min}}, \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}, \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}})$. For each $(s, a) \in S \times A$, let $V_{s,a}$ denote the set of vertices of the local uncertainty polytope $P(s, a)$.

- *States:* $S^{\mathcal{G}}_{\mathsf{max}} = S$ and $S^{\mathcal{G}}_{\mathsf{min}} = S \times A$. Let $S^{\mathcal{G}} = S^{\mathcal{G}}_{\mathsf{max}} \cup S^{\mathcal{G}}_{\mathsf{min}}$.
- *Actions:* $A^{\mathcal{G}}_{\mathsf{max}} = A$ and $A^{\mathcal{G}}_{\mathsf{min}} = \bigcup_{(s,a)\in S\times A} V_{s,a}$. At min-state (s, a) the effective actions are $V_{s,a}$; all remaining min-actions use the implicit min-losing transition. Let $A^{\mathcal{G}} = A^{\mathcal{G}}_{\mathsf{max}} \cup A^{\mathcal{G}}_{\mathsf{min}}$.
- *Transition function* $\delta^{\mathcal{G}} : (S^{\mathcal{G}}_{\mathsf{max}} \times A^{\mathcal{G}}_{\mathsf{max}}) \cup (S^{\mathcal{G}}_{\mathsf{min}} \times A^{\mathcal{G}}_{\mathsf{min}}) \to \Delta(S^{\mathcal{G}})$ is defined as follows:
  - For max-player having state-action pairs $(s, a) \in S^{\mathcal{G}}_{\mathsf{max}} \times A^{\mathcal{G}}_{\mathsf{max}}$, the transition function is given as Dirac distribution with probability mass on state $(s, a) \in S^{\mathcal{G}}_{\mathsf{min}}$
    $$\delta^{\mathcal{G}}(s, a)(s') = \begin{cases} 1 & \text{if } s' = (s, a) \\ 0 & \text{otherwise} \end{cases}$$
  - For an effective min-action $v \in V_{s,a}$ at $(s, a) \in S^{\mathcal{G}}_{\mathsf{min}}$, the transition function is the probability distribution over $S^{\mathcal{G}}_{\mathsf{max}}$ given by
    $$\delta^{\mathcal{G}}((s, a), v)(s') = \begin{cases} v(s') & \text{where } s' \in S^{\mathcal{G}}_{\mathsf{max}} \text{ and } v(s') \text{ denotes the } s'\text{th entry in } v \\ 0 & \text{otherwise} \end{cases}$$
    Every other globally enabled min-action follows the default max-winning transition and is suppressed from the display.
- *Initial state:* $s^{\mathcal{G}}_0 = s_0 \in S^{\mathcal{G}}_{\mathsf{max}}$.
- *Observation functions:* $\mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}$ and $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}$ are functions from states $S^{\mathcal{G}}$ to $\mathcal{O}^{\mathcal{G}}_{\mathsf{max}}$ and $\mathcal{O}^{\mathcal{G}}_{\mathsf{min}}$ respectively.
  - For $s \in S^{\mathcal{G}}_{\mathsf{max}}$, $\mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}(s) = \mathcal{Z}_{\mathsf{a}}(s)$
  - For $(s, a) \in S^{\mathcal{G}}_{\mathsf{min}}$, $\mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}((s, a)) = \mathcal{Z}_{\mathsf{a}}(s)$
  - For $s \in S^{\mathcal{G}}_{\mathsf{max}}$, $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s) = \mathcal{Z}_{\mathsf{e}}(s)$
  - For $(s, a) \in S^{\mathcal{G}}_{\mathsf{min}}$, $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}((s, a)) = \mathcal{Z}_{\mathsf{e}}(s)$

An effective play in $\mathcal{G}$ has the form $\rho = s_0, a_0, s_1, a_1, \ldots \in \mathsf{Plays}_{\mathcal{G}}$, where for every $k \geq 0$, $s_{2k} \in S^{\mathcal{G}}_{\mathsf{max}}$, $a_{2k} \in A^{\mathcal{G}}_{\mathsf{max}}$, $s_{2k+1} = (s_{2k}, a_{2k}) \in S^{\mathcal{G}}_{\mathsf{min}}$, and $a_{2k+1} \in V_{s_{2k},a_{2k}}$. Non-matching min-actions take the implicit min-losing transition. At step t, the history is $h_t = s_0, (o^1_0, o^2_0), s_1, \ldots, s_t, (o^1_t, o^2_t)$, where $o^1_i = \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}(s_i)$ and $o^2_i = \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s_i)$.

A.1 Mapping Plays, Observation Histories and Strategies from $\mathcal{M}$ to $\mathcal{G}$

In this section, we establish a formal mapping between the plays, observation histories and strategies of the RPOMDP $\mathcal{M}$ and the POSG $\mathcal{G}$.

Mapping on Plays:

- Define a mapping $\Theta : \mathsf{Plays}_{\mathcal{M}} \to 2^{\mathsf{Plays}_{\mathcal{G}}}$ as follows. Let $\rho = s_0, a_0, s_1, a_1, \cdots \in \mathsf{Plays}_{\mathcal{M}}$. Then $\Theta(\rho)$ is the set of all plays $\rho' = s'_0, a'_0, s'_1, a'_1, s'_2, a'_2, \cdots \in \mathsf{Plays}_{\mathcal{G}}$ is given index-wise, $\forall k \geq 0 : s'_{2k} = s_k, a'_{2k} = a_k, s'_{2k+1} = (s'_{2k}, a'_{2k}) = (s_k, a_k), a'_{2k+1} = v \in V_{s_k,a_k}$ (arbitrary v). Thus, $\Theta(\rho)$ consists of all plays in G obtained by resolving each transition in ρ using an arbitrary choice of vertex from the uncertainty set at each step. For $X \subseteq \mathsf{Plays}_{\mathcal{M}}$, $\Theta(X) = \bigcup_{\rho \in X} \Theta(\rho)$.
- Define $\Upsilon : \mathsf{Plays}_{\mathcal{G}} \to \mathsf{Plays}_{\mathcal{M}}$ as follows. Let $\rho' = s'_0, a'_0, s'_1, a'_1, s'_2, a'_2, \cdots$ be a play in $\mathcal{G}$. Then $\Upsilon(\rho') = \rho = s_0, a_0, s_1, a_1, \cdots \in \mathsf{Plays}_{\mathcal{M}}$ is given index-wise by $s_i = s'_{2i}$ and $a_i = a'_{2i}$ for every $i \geq 0$. The map Υ is many-to-one. For $Y \subseteq \mathsf{Plays}_{\mathcal{G}}$, define $\Upsilon(Y) = \{\Upsilon(\rho') \mid \rho' \in Y\}$; for $X \subseteq \mathsf{Plays}_{\mathcal{M}}$, define its preimage $\Upsilon^{-1}(X) = \{\rho' \in \mathsf{Plays}_{\mathcal{G}} \mid \Upsilon(\rho') \in X\}$.

Mapping on Observation Histories: $\mathsf{OH}_a, \mathsf{OH}_e$ denote the set of agent and environment observation histories respectively in RPOMDP $\mathcal{M}$. $\mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}, \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$ denote observation histories of $\mathsf{max}, \mathsf{min}$ players respectively in POSG $\mathcal{G}$.

1. Define $\Theta_a : \mathsf{OH}_a \to \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$. For agent observation history $oh^a = o_0, o_1 \cdots, o_t \in \mathsf{OH}_a$, the mapping is given by $\Theta_a(oh^a) = oh^{\mathsf{max}} = o'_0, o'_1, \cdots, o'_{2t}$ is given index-wise, $\forall\, 0 \leq k \leq t-1, o'_{2k} = o'_{2k+1} = o_k, o'_{2t} = o_t$.
2. Define $\Upsilon_a : \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}} \to \mathsf{OH}_a$. For max-player observation history $oh^{\mathsf{max}} = o_0, o_1, \cdots, o_{t-1}, o_t$, the mapping is given by $\Upsilon_a(oh^{\mathsf{max}}) = oh^a = o'_0, o'_1, \cdots, o'_{t/2}$ is given index-wise, $\forall 0 \leq i \leq t/2,\ o'_i = o_{2i}$.
3. Define $\Theta_e : \mathsf{OH}_e \to \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$. Let $oh^e = o_0, o_1, \cdots, o_t \in \mathsf{OH}_e$. The function is given by $\Theta_e(oh^e) = oh^{\mathsf{min}} = o'_0, o'_1, \cdots, o'_{2t}, o'_{2t+1}$ is given index-wise, $\forall 0 \leq k \leq t, o'_{2k} = o'_{2k+1} = o_k$.
4. Define $\Upsilon_e : \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}} \to \mathsf{OH}_e$. Let $oh^{\mathsf{min}} = o_0, o_1, \cdots, o_t \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$. The function is given by $\Upsilon_e(oh^{\mathsf{min}}) = oh^e = o'_0, o'_1, \cdots, o'_{(t-1)/2}$ is given index-wise, $\forall 0 \leq i \leq (t-1)/2,\ o'_i = o_{2i}$.

Lemma A.1. *Let $\star \in \{a, e\}$. $\Theta_\star$ and $\Upsilon_\star$ are inverses of each other.*

Proof.

1. For all $oh = o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$,
$$\Theta_a(\Upsilon_a(oh)) = \Theta_a(o_1, o_2, \cdots, o_{t-1}, o_t) = o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t = oh$$
2. For all $oh = o_1, o_2, \cdots, o_{t-1}, o_t \in \mathsf{OH}_a$,
$$\Upsilon_a(\Theta_a(oh)) = \Upsilon_a(o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t) = o_1, o_2, \cdots, o_{t-1}, o_t = oh$$
3. For all $oh = o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t, o_t \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$,
$$\Theta_e(\Upsilon_e(oh)) = \Theta_e(o_1, o_2, \cdots, o_{t-1}, o_t) = o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t, o_t = oh$$
4. For all $oh = o_1, o_2, \cdots, o_{t-1}, o_t \in \mathsf{OH}_e$,
$$\Upsilon_e(\Theta_e(oh)) = \Upsilon_e(o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t, o_t) = o_1, o_2, \cdots, o_{t-1}, o_t = oh$$

□

Mapping of Strategies For each $(s, a) \in S \times A$, define the barycenter map

$$\mathsf{bar}_{s,a} : \Delta(V_{s,a}) \to P(s, a), \qquad \mathsf{bar}_{s,a}(\lambda) = \sum_{v \in V_{s,a}} \lambda(v) v.$$

For each $P \in P(s, a)$, choose coefficients $\alpha^P_{s,a} \in \Delta(V_{s,a})$ such that $\mathsf{bar}_{s,a}(\alpha^P_{s,a}) = P$. Such coefficients need not be unique; fix one choice for every P once and for all. We identify $\Delta(V_{s,a})$ with distributions over $A^{\mathcal{G}}_{\mathsf{min}}$ supported on $V_{s,a}$. Also fix an arbitrary reference distribution $P^0_{s,a} \in P(s, a)$ for every $(s, a) \in S \times A$.

1. The agent strategy $\sigma \in \Sigma_{\mathcal{M}}$ for RPOMDP $\mathcal{M}$ is given by
$$\sigma : \mathsf{OH}_a \to \bigcup_{o \in \mathcal{O}_\mathsf{a}} \prod_{s:\, \mathcal{Z}_\mathsf{a}(s)=o} \Delta(A)$$
2. The environment strategy $\pi \in \Pi_{\mathcal{M}}$ for RPOMDP $\mathcal{M}$ is given by
$$\pi : \mathsf{OH}_e \times A \to \mathcal{P}.$$
3. The max-player strategy $\sigma' \in \Sigma_{\mathcal{G}}$ for POSG $\mathcal{G}$ from the reduction is given by
$$\sigma' : \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}} \to \bigcup_{o \in \mathcal{O}^{\mathcal{G}}_{\mathsf{max}}} \prod_{s \in S^{\mathcal{G}}_{\mathsf{max}}:\, \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}(s)=o} \Delta(A^{\mathcal{G}}_{\mathsf{max}})$$
(recall that max and min states alternate)
4. The min-player strategy $\pi' \in \Pi_{\mathcal{G}}$ for POSG $\mathcal{G}$ from the reduction is given by
$$\pi' : \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}} \to \bigcup_{o \in \mathcal{O}^{\mathcal{G}}_{\mathsf{min}}} \prod_{s \in S^{\mathcal{G}}_{\mathsf{min}}:\, \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s)=o} \Delta(A^{\mathcal{G}}_{\mathsf{min}})$$
(recall that max and min states alternate)

Define the mappings

- $\mathfrak{C}_a : \Sigma_{\mathcal{M}} \to \Sigma_{\mathcal{G}}$. For all $\sigma \in \Sigma_{\mathcal{M}}$ and $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$,
$$\mathfrak{C}_a(\sigma)(oh) = \sigma(\Upsilon_a(oh)).$$
- $\mathfrak{D}_a : \Sigma_{\mathcal{G}} \to \Sigma_{\mathcal{M}}$. For all $\sigma' \in \Sigma_{\mathcal{G}}$ and $oh \in \mathsf{OH}_a$,
$$\mathfrak{D}_a(\sigma')(oh) = \sigma'(\Theta_a(oh)).$$
- $\mathfrak{C}_e : \Pi_{\mathcal{M}} \to \Pi_{\mathcal{G}}$. For all $\pi \in \Pi_{\mathcal{M}}$, $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$, and $(s,a) \in S^{\mathcal{G}}_{\mathsf{min}}$ with $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}((s,a)) = \mathrm{last}(oh)$, use the component notation above.
$$\mathfrak{C}_e(\pi)(oh)((s,a)) = \alpha^{\pi(\Upsilon_e(oh),a)(s,a)}_{s,a}.$$
- $\mathfrak{D}_e : \Pi_{\mathcal{G}} \to \Pi_{\mathcal{M}}$. For all $\pi' \in \Pi_{\mathcal{G}}$, $oh \in \mathsf{OH}_e$, and $a \in A$, define the complete assignment $\mathfrak{D}_e(\pi')(oh,a) \in \mathcal{P}$ componentwise as follows.
$$\mathfrak{D}_e(\pi')(oh,a)(x,b) = \begin{cases} \mathsf{bar}_{x,a}(\pi'(\Theta_e(oh))((x,a))) & \text{if } b = a \text{ and } \mathcal{Z}_\mathsf{e}(x) = \mathrm{last}(oh), \\ P^0_{x,b} & \text{otherwise.} \end{cases}$$
Hence every component belongs to its corresponding local uncertainty set.

Remark A.2. *For every $\pi \in \Pi_{\mathcal{M}}$, $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$, and compatible (s,a),*
$$\pi(\Upsilon_e(oh),a)(s,a) = \mathsf{bar}_{s,a}(\mathfrak{C}_e(\pi)(oh)((s,a)))\,.$$

Lemma A.3. *The agent maps $\mathfrak{C}_a$ and $\mathfrak{D}_a$ are mutual inverses. The environment maps satisfy, at every compatible oh, s, and a,*
$$\mathfrak{D}_e(\mathfrak{C}_e(\pi))(oh,a)(s,a) = \pi(oh,a)(s,a),$$
$$\mathsf{bar}_{s,a}(\mathfrak{C}_e(\mathfrak{D}_e(\pi'))(oh)((s,a))) = \mathsf{bar}_{s,a}(\pi'(oh)((s,a)))\,.$$
Thus both compositions preserve the next-state distribution.

Proof. The agent identities follow from Lemma A.1. The environment identities follow from $\mathsf{bar}_{s,a}(\alpha^P_{s,a}) = P$ and the two history identities in Lemma A.1.

□

A.2 Lifting Objectives

Let W be a parity objective in RPOMDP $\mathcal{M}$ given by the priority function $p : S \to \{0, 1, \cdots, d\},\ d \in \mathbb{N}$. Then, the *lifted* objective in POSG $\mathcal{G}$, $W' \subseteq \mathsf{Plays}^{\mathcal{G}}$ is given the priority function $p^{\mathcal{G}} : S^{\mathcal{G}}_{\mathsf{max}} \cup S^{\mathcal{G}}_{\mathsf{min}} \to \{0, 1, \cdots, d\}$ defined as

$$p^{\mathcal{G}}(s) = \begin{cases} p(s) & \text{if } s \in S^{\mathcal{G}}_{\mathsf{max}} \\ p(s') & \text{if } s = (s', a) \in S^{\mathcal{G}}_{\mathsf{min}} \end{cases}$$

Lemma A.4. *Let W be parity objective in RPOMDP $\mathcal{M}$ given by the priority function p and let W' be the* lifted *objective in POSG $\mathcal{G}$. Then,*

$$\rho \in W \iff \Theta(\rho) \subseteq W'$$

Proof. We have $W = \mathrm{Parity}_{\mathcal{M}}(p)$ and $W' = \mathrm{Parity}_{\mathcal{G}}\big(p^{\mathcal{G}}\big)$ where $p^{\mathcal{G}}$ is defined as above. Any play $\rho = s_0, a_0, s_1, a_1, \cdots \in \mathsf{Plays}_{\mathcal{M}}$ has parity sequence: $p(\rho) = p_0, p_1, p_2, \cdots$ such that $p_i = p(s_i)$. Any play in $\Theta(\rho)$ is of the form $\rho' = s_0, a_0, (s_0, a_0), a'_0, s_1, \cdots$ with parity sequence $p^{\mathcal{G}}(\rho') = p_0, p_0, p_1, p_1, \cdots$, since $p^{\mathcal{G}}(s_i) = p^{\mathcal{G}}((s_i, a_i)) = p(s_i)$. Due to repeating values $\liminf_{i\to\infty}$ of the sequence do not change, then $\liminf_{i\to\infty}(p(\rho)) = \liminf_{i\to\infty}(p^{\mathcal{G}}(\rho'))$. Thus, $\rho \in W \implies \rho' \in W' \iff \Theta(\rho) \subseteq W'$. The reverse direction follows similarly. □

A.3 Value Preservation

Lemma A.5. *The strategy maps preserve the induced probability measure in both directions. In particular,*

$$\mathbb{P}^{\sigma,\pi}_{\mathcal{M}}(W) = \mathbb{P}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}_{\mathcal{G}}(W') \qquad \text{for all } \sigma \in \Sigma_{\mathcal{M}},\ \pi \in \Pi_{\mathcal{M}},$$

$$\mathbb{P}^{\sigma',\pi'}_{\mathcal{G}}(W') = \mathbb{P}^{\mathfrak{D}_a(\sigma'),\mathfrak{D}_e(\pi')}_{\mathcal{M}}(W) \qquad \text{for all } \sigma' \in \Sigma_{\mathcal{G}},\ \pi' \in \Pi_{\mathcal{G}},$$

where $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ is a parity objective of $\mathcal{M}$ and $W' = \Theta(W)$ is the corresponding objective in $\mathcal{G}$.

Proof. For a history h, let $\mathrm{Cyl}(h)$ denote the set of all runs having h as a prefix. Since the probability measure over runs is uniquely determined by its values on cylinder sets, it suffices to show that corresponding cylinder sets have the same probability.

Let $H_{\mathcal{M}}$(resp. $H_{\mathcal{G}}$) denote the set of histories in RPOMDP $\mathcal{M}$ (resp. POSG $\mathcal{G}$). Define the mapping $\Theta_h : H_{\mathcal{M}} \to 2^{H_{\mathcal{G}}}$ as follows: Let $h = s_0, (o^a_0, o^e_0), s_1, \ldots s_t, (o^a_t, o^e_t) \in H_{\mathcal{M}}$. Then $\Theta_h(h)$ is the set of all histories $h' = s'_0, (o^{max}_0, o^{min}_0), s'_1, \ldots s'_{2t}, (o^{max}_{2t}, o^{min}_{2t})$ is given index-wise, $\forall k \geq 0 : s'_{2k} = s_k, s'_{2k+1} = (s_k, a_k)$ where $a_k \in A^{\mathcal{G}}_{\mathsf{max}}$ and $\forall i \geq 0\ o^{max}_i = \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}(s'_i), o^{min}_i = \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s'_i)$. Thus, $\Theta_h(h)$ consists of all histories in G obtained by resolving each transition in h using an arbitrary choice of action from $A^{\mathcal{G}}_{\mathsf{max}}$ at each step.

We show that for a finite history h in W, there exists corresponding set of histories $H' = \Theta_h(h)$ in W' such that

$$\mathbb{P}^{\sigma,\pi}_{\mathcal{M}}(\mathrm{Cyl}(h)) = \sum_{h' \in H'} \mathbb{P}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}_{\mathcal{G}}(\mathrm{Cyl}(h'))$$

The proof follows induction on the length of h with base condition $h = s_0$ and $H' = \{s_0\}$ with values $\mathbb{P}^{\sigma,\pi}_{\mathcal{M}}(\mathrm{Cyl}(h)) = \sum_{h' \in H'} \mathbb{P}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}_{\mathcal{G}}(\mathrm{Cyl}(h')) = 1$.

For induction, let $h = h_1, s_1, (o_1, o'_1)$ where $o_1 = \mathcal{Z}_{\mathsf{a}}(s_1), o'_1 = \mathcal{Z}_{\mathsf{e}}(s_1)$ and h_1 ending in state s with observations history pair (o, o'). Let $H'_1 = \Theta_h(h_1)$. A history $h' \in H'$ is of the form $h' = h'_1, (s, a), (o, o'), s_1, (o_1, o'_1)$ where $h'_1 \in H'_1$, $a \in A$(an arbitrary action a). Let oh^a, oh^e be the agent and environment observations histories for history h_1. Let oh^{max} be max player observation history for all histories $h'_1 \in H'_1$. Then $oh^a = \Upsilon_a(oh^{\mathsf{max}})$. Let oh^{min} be min player observation history for $h'_1, (s, a), (o, o')$ where $h' \in H'_1$. Then $oh^e = \Upsilon_e(oh^{\mathsf{min}})$.

For a given $\sigma' \in \Sigma_{\mathcal{G}}, \pi' \in \Pi_{\mathcal{G}}$ in history h', the transition probability from s to (s, a) is given by $\sigma'(oh^{\mathsf{max}})(s)(a)$ and the transition probability from (s, a) to s_1 is given by all the transitions on the vertices $v \in V_{s,a}$ i.e., $\Big(\sum_{v \in V_{s,a}} \pi'(oh^{\mathsf{min}})((s, a))(v) \cdot v(s_1)\Big)$.

$$\sum_{h' \in H'} \mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}(\mathrm{Cyl}(h'))$$

$$= \sum_{h' \in H'} \left(\mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}(\mathrm{Cyl}(h'_1)) \times \mathfrak{C}_a(\sigma)(oh^{\mathsf{max}})(s)(a) \times \left(\sum_{v \in V_{s,a}} \mathfrak{C}_e(\pi)(oh^{\mathsf{min}})((s,a))(v) \cdot v(s_1) \right) \right)$$

$$= \sum_{h' \in H'_1} \mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}(\mathrm{Cyl}(h'_1)) \times \sum_{a \in A} \left(\mathfrak{C}_a(\sigma)(oh^{\mathsf{max}})(s)(a) \times \left(\sum_{v \in V_{s,a}} \mathfrak{C}_e(\pi)(oh^{\mathsf{min}})((s,a))(v) \cdot v(s_1) \right) \right)$$

By induction hypothesis, $\sum_{h' \in H'_1} \mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}(\mathrm{Cyl}(h'_1)) = \mathbb{P}_{\mathcal{M}}^{\sigma,\pi}(\mathrm{Cyl}(h_1))$. From definition of $\mathfrak{C}_a$, $\mathfrak{C}_a(\sigma)(oh^{\mathsf{max}})(s)(a) = \sigma(\Upsilon_a(oh^{\mathsf{max}}))(s)(a) = \sigma(oh^a)(s)(a)$. From remark A.2, $\left(\sum_{v \in V_{s,a}} \mathfrak{C}_e(\pi)(oh^{\mathsf{min}})((s,a))(v) \cdot v(s_1)\right) = \pi(\Upsilon_e(oh^{\mathsf{min}}), a)(s,a)(s_1) = \pi(oh^e, a)(s,a)(s_1)$. So,

$$\sum_{h' \in H'} \mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}(\mathrm{Cyl}(h')) = \mathbb{P}_{\mathcal{M}}^{\sigma,\pi}(\mathrm{Cyl}(h_1)) \times \sum_{a \in A} (\sigma(oh^a)(s)(a) \times \pi(oh^e, a)(s,a)(s_1))$$
$$= \mathbb{P}_{\mathcal{M}}^{\sigma,\pi}(\mathrm{Cyl}(h))$$

The converse cylinder identity follows by the same induction, replacing the vertex distribution by its barycenter. The second identity of Lemma A.3 shows that this replacement preserves every next-state probability. Equality on all cylinders yields both asserted measure equalities and hence both objective-probability equalities. □

Lemma A.6. *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and let $W' = \Theta(W)$ be the corresponding objective in $\mathcal{G}$. Then,*

$$\mathrm{Val}_{\mathcal{M}}(W) = \mathrm{Val}_{\mathcal{G}}(W').$$

Proof. For fixed $\sigma \in \Sigma_{\mathcal{M}}$, set

$$q_{\mathcal{M}}(\sigma) = \inf_{\pi \in \Pi_{\mathcal{M}}} \mathbb{P}_{\mathcal{M}}^{\sigma,\pi}(W), \qquad q_{\mathcal{G}}(\sigma') = \inf_{\pi' \in \Pi_{\mathcal{G}}} \mathbb{P}_{\mathcal{G}}^{\sigma',\pi'}(W').$$

The first equality in Lemma A.5 gives $q_{\mathcal{G}}(\mathfrak{C}_a(\sigma)) \leq q_{\mathcal{M}}(\sigma)$, because every $\pi \in \Pi_{\mathcal{M}}$ has the image $\mathfrak{C}_e(\pi) \in \Pi_{\mathcal{G}}$. Conversely, for every $\pi' \in \Pi_{\mathcal{G}}$, the second equality in Lemma A.5 and $\mathfrak{D}_a(\mathfrak{C}_a(\sigma)) = \sigma$ give

$$\mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\pi'}(W') = \mathbb{P}_{\mathcal{M}}^{\sigma,\mathfrak{D}_e(\pi')}(W),$$

hence $q_{\mathcal{M}}(\sigma) \leq q_{\mathcal{G}}(\mathfrak{C}_a(\sigma))$. Therefore

$$q_{\mathcal{M}}(\sigma) = q_{\mathcal{G}}(\mathfrak{C}_a(\sigma)) \quad \text{for every } \sigma \in \Sigma_{\mathcal{M}}.$$

Finally, $\mathfrak{C}_a$ is a bijection by Lemma A.3, so taking the outer supremum yields

$$\mathrm{Val}_{\mathcal{M}}(W) = \sup_{\sigma \in \Sigma_{\mathcal{M}}} q_{\mathcal{M}}(\sigma) = \sup_{\sigma \in \Sigma_{\mathcal{M}}} q_{\mathcal{G}}(\mathfrak{C}_a(\sigma))$$
$$= \sup_{\sigma' \in \Sigma_{\mathcal{G}}} q_{\mathcal{G}}(\sigma') = \mathrm{Val}_{\mathcal{G}}(W').$$

□

Lemma A.7. *For mapped strategies, projection preserves the supported plays in both directions:*

$$\Upsilon\left(\mathsf{Plays}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}\right) = \mathsf{Plays}_{\mathcal{M}}^{\sigma,\pi},$$
$$\Upsilon\left(\mathsf{Plays}_{\mathcal{G}}^{\sigma',\pi'}\right) = \mathsf{Plays}_{\mathcal{M}}^{\mathfrak{D}_a(\sigma'),\mathfrak{D}_e(\pi')}.$$

Proof. We prove the first equality by induction on finite effective prefixes; the second follows by the same argument using the reverse maps. The initial prefixes coincide. Suppose corresponding prefixes end in the RPOMDP state s and the POSG max-state s. Their agent observation histories correspond under Θ_a and Υ_a, and

$$\mathfrak{C}_a(\sigma)(\Theta_a(oh^a))(s) = \sigma(oh^a)(s).$$

Consequently, an action a has positive probability after one prefix exactly when it has positive probability after the other. The POSG then moves deterministically to (s, a).

At the min-state (s, a), let $\lambda = \mathfrak{C}_e(\pi)(\Theta_e(oh^e))((s, a))$ and $P = \pi(oh^e, a)(s, a) = \mathsf{bar}_{s,a}(\lambda)$. For every successor s', nonnegativity gives

$$P(s') > 0 \iff \lambda(v)v(s') > 0 \text{ for some } v \in V_{s,a}.$$

Thus an RPOMDP extension to s' is supported exactly when some POSG extension through an effective vertex v to s' is supported. This proves the induction step and the first projected-prefix equality. For arbitrary σ', π', replace λ by $\pi'(oh)((s, a))$ and use the barycenter identity in Lemma A.3; the same induction proves the second equality. Finally, an infinite play is supported exactly when all its finite prefixes are supported, yielding both displayed play-set equalities.

□

Lemma A.8 (Sure winning analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and $W' = \Upsilon^{-1}(W) = \Theta(W)$ its lifted objective in $\mathcal{G}$. The preimage Υ^{-1} was defined in Appendix A.1. For $\sigma \in \Sigma_{\mathcal{M}}$, let $\sigma' = \mathfrak{C}_a(\sigma)$. Then,*

$$\sigma \text{ is sure winning for } W \iff \sigma' \text{ is sure winning for } W'.$$

Proof. Suppose first that σ is sure winning and take arbitrary $\pi' \in \Pi_{\mathcal{G}}$. By Lemma A.7,

$$\Upsilon\left(\mathsf{Plays}_{\mathcal{G}}^{\sigma',\pi'}\right) = \mathsf{Plays}_{\mathcal{M}}^{\sigma,\mathfrak{D}_e(\pi')} \subseteq W,$$

so $\mathsf{Plays}_{\mathcal{G}}^{\sigma',\pi'} \subseteq \Upsilon^{-1}(W) = W'$. Conversely, suppose that σ' is sure winning and take arbitrary $\pi \in \Pi_{\mathcal{M}}$. The first equality in Lemma A.7 gives $\Upsilon(\mathsf{Plays}_{\mathcal{G}}^{\sigma',\mathfrak{C}_e(\pi)}) = \mathsf{Plays}_{\mathcal{M}}^{\sigma,\pi}$; since the game plays lie in W', their projections lie in W. □

Lemma A.9 (Almost-sure analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and $W' = \Theta(W)$ its lifted objective in $\mathcal{G}$. For $\sigma \in \Sigma_{\mathcal{M}}$, let $\sigma' = \mathfrak{C}_a(\sigma)$. Then,*

$$\sigma \text{ is almost-sure winning for } W \iff \sigma' \text{ is almost-sure winning for } W'.$$

Proof. The fixed-strategy equality established in the proof of Lemma A.6 gives

$$\inf_{\pi \in \Pi_{\mathcal{M}}} \mathbb{P}_{\mathcal{M}}^{\sigma,\pi}(W) = \inf_{\pi' \in \Pi_{\mathcal{G}}} \mathbb{P}_{\mathcal{G}}^{\sigma',\pi'}(W').$$

One side equals 1 if and only if the other does. □

Lemma A.10 (Limit-sure Analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and $W' = \Theta(W)$ is the corresponding objective in $\mathcal{G}$. Then,*

$$W \text{ is limit-sure winning} \iff W' \text{ is limit-sure winning}$$

Proof. Follows directly from lemma A.6. □

Lemma A.11 (Quantitative Analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and $W' = \Theta(W)$ is the corresponding objective in $\mathcal{G}$. Then, for $k \in (0, 1)$*

$$\mathrm{Val}_{\mathcal{M}}(W) \geq k \iff \mathrm{Val}_{\mathcal{G}}(W') \geq k$$

Proof. Follows directly from lemma A.6. □

A.4 Proof of Theorem 3.1

Proof. Because parity subsumes all ω-regular objectives, from Lemma A.6, $\mathrm{Val}_{\mathcal{M}}(W) = \mathrm{Val}_{\mathcal{G}}(W')$ for any ω-regular objective W in $\mathcal{M}$ and its lifted objective W' in $\mathcal{G}$.
The "consequently" statement follows from lemmas A.8, A.9, A.10, and A.11. □

B The Pre-min-transformation from ac-POSG $\mathcal{G}$ to $\mathcal{G}'$

In this section, we begin with the transformation from $\mathcal{G}$ to $\mathcal{G}'$.

Consider any *alternating control* POSG $\mathcal{G}$. In $\mathcal{G}$ the max-player selects an action from a max-state and the output is immediately a min-state (and vice versa). To make the subsequent RPOMDP construction cleaner, we first insert a dummy max-state d_s before every min-state $s \in S_{\mathsf{min}}$, so that the max-player always acts *twice* in a row before the min-player acts. The max-action at a dummy state is always the fixed action $\top$ with a deterministic transition to the corresponding min-state; it therefore carries no strategic content.

Definition B.1 (Pre-min transformation)**.** Given an *alternating control* POSG $\mathcal{G} = (S_{\mathsf{max}}, S_{\mathsf{min}}, A_{\mathsf{max}}, A_{\mathsf{min}}, \delta, s_0, \mathcal{O}_{\mathsf{max}}, \mathcal{O}_{\mathsf{min}}, \mathcal{Z}_{\mathsf{max}}, \mathcal{Z}_{\mathsf{min}})$, the *Pre*-min *transformation* of $\mathcal{G}$ is the POSG $\mathcal{G}' = (S'_{\mathsf{max}}, S'_{\mathsf{min}}, A'_{\mathsf{max}}, A'_{\mathsf{min}}, \delta', s_0, \mathcal{O}'_{\mathsf{max}}, \mathcal{O}'_{\mathsf{min}}, \mathcal{Z}'_{\mathsf{max}}, \mathcal{Z}'_{\mathsf{min}})$ where:

- *State spaces:*

$$S'_{\mathsf{max}} \;=\; S_{\mathsf{max}} \;\cup\; \underbrace{\{\, d_s \mid s \in S_{\mathsf{min}} \,\}}_{D}, \qquad S'_{\mathsf{min}} \;=\; S_{\mathsf{min}},$$

  where D is a set of fresh dummy max-states, one for each min-state.

- *Actions:*

$$A'_{\mathsf{max}} \;=\; A_{\mathsf{max}} \;\cup\; \{\top\}, \qquad A'_{\mathsf{min}} \;=\; A_{\mathsf{min}},$$

  where $\top \notin A_{\mathsf{max}}$ is fresh. All actions in A'_{max} are enabled at every max-state; A_{max} is effective at original max-states and $\top$ is effective at dummy states.

- *Transition function δ':*
  - For $s_1 \in S_{\mathsf{max}}$, $a_1 \in A_{\mathsf{max}}$, and $s_2 \in S_{\mathsf{min}}$: $\delta'(s_1, a_1)(d_{s_2}) = \delta(s_1, a_1)(s_2)$.
  - For $s_2 \in S_{\mathsf{min}}$ and $a_2 \in A_{\mathsf{min}}$: $\delta'(s_2, a_2) = \delta(s_2, a_2)$.
  - For $d_s \in D$ and action $\top$: $\delta'(d_s, \top) = \mathrm{Dirac}(s)$.
  - Every other max state-action pair follows the implicit max-losing transition.

- *Observations:*

$$\mathcal{O}'_{\mathsf{max}} \;=\; \mathcal{O}_{\mathsf{max}} \;\cup\; \underbrace{\{\, o^{\dagger}_x \mid x \in \mathcal{O}_{\mathsf{max}} \,\}}_{\mathcal{O}^{\dagger}}, \qquad \mathcal{O}'_{\mathsf{min}} \;=\; \mathcal{O}_{\mathsf{min}} \;\cup\; \underbrace{\{\, o^{\#}_x \mid x \in \mathcal{O}_{\mathsf{min}} \,\}}_{\mathcal{O}^{\#}}.$$

- *Observation functions:*
  - $\forall s \in S_{\mathsf{max}} \cup S_{\mathsf{min}}$: $\mathcal{Z}'_{\mathsf{max}}(s) = \mathcal{Z}_{\mathsf{max}}(s)$ and $\mathcal{Z}'_{\mathsf{min}}(s) = \mathcal{Z}_{\mathsf{min}}(s)$.
  - $\forall d_s \in D$: $\mathcal{Z}'_{\mathsf{max}}(d_s) = o^{\dagger}_{\mathcal{Z}_{\mathsf{max}}(s)}$ and $\mathcal{Z}'_{\mathsf{min}}(d_s) = o^{\#}_{\mathcal{Z}_{\mathsf{min}}(s)}$.

B.1 Mapping plays, observation histories and strategies from $\mathcal{G}$ to $\mathcal{G}'$

Mapping on plays. We define a map $\Gamma : \mathsf{Plays}_{\mathcal{G}} \to \mathsf{Plays}_{\mathcal{G}'}$. Let $h = s_0, a_0, s_1, a_1, s_2, a_2, \cdots \in \mathsf{Plays}_{\mathcal{G}}$. Since $\mathcal{G}$ is alternating control, even-indexed states lie in S_{max} and odd-indexed states lie in S_{min}. We set $\Gamma(h) = h' = s'_0, a'_0, s'_1, a'_1, \cdots$ index-wise as: For $k \geq 0$,

$$\begin{aligned} s'_{3k} &= s_{2k} \\ a'_{3k} &= a_{2k} \\ s'_{3k+1} &= d_{s_{2k+1}} \\ a'_{3k+1} &= \top \\ s'_{3k+2} &= s_{2k+1} \\ a'_{3k+2} &= a_{2k+1} \end{aligned}$$

For $X \subseteq \mathsf{Plays}_{\mathcal{G}}$, define $\Gamma(X) = \{\Gamma(\rho) \mid \rho \in X\}$.

Observation B.2. *By construction of $\mathcal{G}'$, $\Gamma(h) \in \mathsf{Plays}_{\mathcal{G}'}$ for every $h \in \mathsf{Plays}_{\mathcal{G}}$, so Γ is well-defined.*

Observation B.3. *Every play $h' = s_1, a_1, s_2, a_2, \cdots \in \mathsf{Plays}_{\mathcal{G}'}$ satisfies, for each $k \geq 0$:*

$$s_{3k+1} \in S_{\mathsf{max}}, \quad a_{3k+1} \in A_{\mathsf{max}}, \quad s_{3k+2} \in D, \quad a_{3k+2} = \top, \quad s_{3k+3} \in S_{\mathsf{min}}, \quad a_{3k+3} \in A_{\mathsf{min}}$$

Lemma B.4. *Let $h' \in \mathsf{Plays}_{\mathcal{G}'}$ and let h be obtained by removing all occurrences of states in D and the action $\top$ from h'. Then $h \in \mathsf{Plays}_{\mathcal{G}}$ and $\Gamma(h) = h'$.*

Lemma B.5. *The map $\Gamma : \mathsf{Plays}_{\mathcal{G}} \to \mathsf{Plays}_{\mathcal{G}'}$ is a bijection.*

Mapping on observation histories. We extend Γ to observation histories. For the max-player, define $\Gamma^{\mathsf{max}}_{\mathcal{O}} : \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}} \to \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$ as follows. Let $oh = x_0, y_0, \ldots, x_{n-1}, y_{n-1}, x_n \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$. Then $\Gamma^{\mathsf{max}}_{\mathcal{O}}(oh) = x_0, z_0, y_0, x_1, z_1, y_1, \ldots, x_{n-1}, z_{n-1}, y_{n-1}, x_n$ where each $z_i = o^{\dagger}_{y_i}$.

Since $\Gamma^{\mathsf{max}}_{\mathcal{O}}(oh)$ cannot end with a $o^{\dagger}$-observation by construction, the map is not surjective. We nevertheless define an inverse $(\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1} : \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}} \to \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$ by removing all occurrences of observations in $\mathcal{O}^{\dagger}$. We define $\Gamma^{\mathsf{min}}_{\mathcal{O}}$ and $(\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}$ for the min-player analogously.

Lemma B.6. *Let $\star \in \{\mathsf{max}, \mathsf{min}\}$. For all $oh \in \mathsf{OH}^{\mathcal{G}}_{\star}$, $(\Gamma^{\star}_{\mathcal{O}})^{-1}(\Gamma^{\star}_{\mathcal{O}}(oh)) = oh$.*

Lemma B.7. *Let $\star \in \{\mathsf{max}, \mathsf{min}\}$ and let $o^{\star} \in \mathcal{O}^{\dagger}$ if $\star = \mathsf{max}$ and $o^{\star} \in \mathcal{O}^{\#}$ otherwise. For all $oh' = (oh'_1, o) \in \mathsf{OH}^{\mathcal{G}'}_{\star}$,*

$$\Gamma^{\star}_{\mathcal{O}}\big((\Gamma^{\star}_{\mathcal{O}})^{-1}(oh'_1, o)\big) = \begin{cases} oh' & \text{if } o \neq o^{\star}, \\ oh'_1 & \text{otherwise.} \end{cases}$$

Mapping of strategies. As in the main text, each player selects a complete prescription based only on its observation history. Thus, for $\star \in \{\mathsf{max}, \mathsf{min}\}$, a strategy in $\mathcal{G}$ has type

$$\sigma_{\star} : \mathsf{OH}^{\mathcal{G}}_{\star} \to \bigcup_{o \in \mathcal{O}_{\star}} \prod_{s \in S_{\star} :\, \mathcal{Z}_{\star}(s) = o} \Delta(A_{\star}),$$

and the corresponding strategies in $\mathcal{G}'$ use the same direct product with $S'_{\mathsf{max}}, S'_{\mathsf{min}}, A'_{\mathsf{max}}, A'_{\mathsf{min}}$. The expressions below evaluate the selected assignment at a compatible state; that state is not an input to the strategy.

We define the following four maps lifting strategies between $\mathcal{G}$ and $\mathcal{G}'$:

1. $\Phi_{\mathsf{max}} : \Sigma_{\mathcal{G}} \to \Sigma_{\mathcal{G}'}$: for each $\sigma \in \Sigma_{\mathcal{G}}$, $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$ and $\forall s \in S_{\mathsf{max}} \cup D$ such that $\mathcal{Z}'_{\mathsf{max}}(s) = \mathrm{last}(oh') = \mathrm{last}((\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh'))$, then
$$\Phi_{\mathsf{max}}(\sigma)(oh')(s) = \sigma\left((\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh')\right)(s)$$
If $\mathcal{Z}'_{\mathsf{max}}(s) = \mathrm{last}(oh') \in \mathcal{O}^{\dagger}$, then $\Phi_{\mathsf{max}}(\sigma)(oh')(s) = \mathrm{Dirac}(\top)$
2. $\Phi_{\mathsf{min}} : \Pi_{\mathcal{G}} \to \Pi_{\mathcal{G}'}$: for each $\pi \in \Pi_{\mathcal{G}}$, $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}}$ and $\forall s \in S_{\mathsf{min}}$ such that $\mathcal{Z}'_{\mathsf{min}}(s) = \mathrm{last}(oh')$, then
$$\Phi_{\mathsf{min}}(\pi)(oh')(s) = \pi\big((\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh')\big)(s)$$
3. $\Psi_{\mathsf{max}} : \Sigma_{\mathcal{G}'} \to \Sigma_{\mathcal{G}}$: for each $\sigma' \in \Sigma_{\mathcal{G}'}$, $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$ and $\forall s \in S_{\mathsf{max}}$ such that $\mathcal{Z}_{\mathsf{max}}(s) = \mathrm{last}(oh)$, then
$$\Psi_{\mathsf{max}}(\sigma')(oh)(s) = \sigma'(\Gamma^{\mathsf{max}}_{\mathcal{O}}(oh))(s).$$
4. $\Psi_{\mathsf{min}} : \Pi_{\mathcal{G}'} \to \Pi_{\mathcal{G}}$: for each $\pi' \in \Pi_{\mathcal{G}'}$, $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$ and $\forall s \in S_{\mathsf{min}}$ such that $\mathcal{Z}_{\mathsf{min}}(s) = \mathrm{last}(oh)$, then
$$\Psi_{\mathsf{min}}(\pi')(oh)(s) = \pi'\big(\Gamma^{\mathsf{min}}_{\mathcal{O}}(oh)\big)(s).$$

Lemma B.8. *For $\star \in \{\mathsf{max}, \mathsf{min}\}$, $\Phi_{\star}$ and $\Psi_{\star}$ are inverses of each other: $\Phi_{\star} = (\Psi_{\star})^{-1}$ and $\Psi_{\star} = (\Phi_{\star})^{-1}$.*

Proof. Without loss of generality, let $\star = \mathsf{min}$. We first show $\Psi_{\mathsf{min}}(\Phi_{\mathsf{min}}(\pi))(oh)(s) = \pi(oh)(s)$ for all $\pi \in \Pi_{\mathcal{G}}$, $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$ and $s \in S_{\mathsf{min}}$ such that $\mathcal{Z}_{\mathsf{min}}(s) = \mathrm{last}(oh)$.

$$\begin{aligned}
\Psi_{\mathsf{min}}(\Phi_{\mathsf{min}}(\pi))(oh)(s) &= \Phi_{\mathsf{min}}(\pi)\big(\Gamma^{\mathsf{min}}_{\mathcal{O}}(oh)\big)(s) && \text{by definition of } \Psi_{\mathsf{min}} \\
&= \pi\big((\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}\big(\Gamma^{\mathsf{min}}_{\mathcal{O}}(oh)\big)\big)(s) && \text{by definition of } \Phi_{\mathsf{min}} \\
&= \pi(oh)(s) && \text{by Lemma B.6.}
\end{aligned}$$

Conversely, for all $\pi' \in \Pi_{\mathcal{G}'}$, $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}}$ and $s \in S_{\mathsf{min}}$ such that $\mathcal{Z}'_{\mathsf{min}}(s) = \mathrm{last}(oh')$:

$$\begin{aligned}
\Phi_{\mathsf{min}}(\Psi_{\mathsf{min}}(\pi'))(oh')(s) &= \Psi_{\mathsf{min}}(\pi')\big((\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh')\big)\,(s) && \text{by definition of } \Phi_{\mathsf{min}} \\
&= \pi'\big(\Gamma^{\mathsf{min}}_{\mathcal{O}}\big((\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh')\big)\big)\,(s) && \text{by definition of } \Psi_{\mathsf{min}} \\
&= \pi'(oh')(s) && \text{by Lemma B.7 and } \mathrm{last}(oh') \notin \mathcal{O}^{\#}.
\end{aligned}$$

□

B.2 Lifting objectives from $\mathcal{G}$ to $\mathcal{G}'$

Let $p : S_{\mathsf{max}} \cup S_{\mathsf{min}} \to \{0, 1, \ldots, d\}$ be the priority function for $\mathcal{G}$. Define $p' : S'_{\mathsf{max}} \cup S'_{\mathsf{min}} \to \{0, 1, \ldots, d\}$ by $p'(s) = p(s)$ for $s \in S_{\mathsf{max}} \cup S_{\mathsf{min}}$ and $p'(d_s) = p(s)$ for $d_s \in D$.

Lemma B.9. *Let $\mathcal{G}$, $\mathcal{G}'$, p, and p' be as above, and let $\rho \in \mathsf{Plays}_{\mathcal{G}}$. Then*

$$\rho \in \mathrm{Parity}_{\mathcal{G}}(p) \iff \Gamma(\rho) \in \mathrm{Parity}_{\mathcal{G}'}(p')\,.$$

Proof. The play $\Gamma(\rho)$ visits every dummy state $d_s \in D$ exactly once between visiting $s \in S_{\mathsf{max}}$ and $s \in S_{\mathsf{min}}$; $p'(d_s) = p(s)$, so no new priority value is introduced. The sequence of priorities seen along $\Gamma(\rho)$ is therefore a repetition of the sequence seen along ρ (each value appears twice in a row at the dummy-state interleaving), which does not change the $\liminf$. Hence $\Gamma(\rho)$ satisfies the parity condition with respect to p' if and only if ρ satisfies it with respect to p. □

B.3 Value Preservation

Lemma B.10. *The strategy maps preserve objective probabilities in both directions:*

$$\begin{aligned}
\mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W) &= \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(W'), && \text{for all } \sigma \in \Sigma_{\mathcal{G}},\ \pi \in \Pi_{\mathcal{G}}, \\
\mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(W') &= \mathbb{P}^{\Psi_{\mathsf{max}}(\sigma'),\Psi_{\mathsf{min}}(\pi')}_{\mathcal{G}}(W), && \text{for all } \sigma' \in \Sigma_{\mathcal{G}'},\ \pi' \in \Pi_{\mathcal{G}'},
\end{aligned}$$

where $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ is a parity objective of $\mathcal{G}$ and $W' = \Gamma(W)$ is its corresponding objective in $\mathcal{G}'$.

Proof. Let $\rho \in W$ be a play of $\mathcal{G}$ and let $\rho' = \Gamma(\rho) \in W'$. It suffices to show that for every finite history h in W the corresponding history h' (constructed index-wise exactly as Γ constructs ρ' from ρ, but truncated to end at the same state as h) satisfies

$$\mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h)) = \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h')).$$

Since the probability measure over plays is uniquely determined by its values on cylinder sets, this suffices. The proof proceeds by induction on $|h|$. The base case $h = s_0$ holds trivially since $h' = s_0$ and both cylinders have probability 1.

For the inductive step, write $h = h_1, s$ and distinguish two cases.

Case 1: $s \in S_{\mathsf{max}}$**.** Then $\mathrm{last}(h_1) \in S_{\mathsf{min}}$. By construction, $h' = h'_1, s$ where h'_1 is the prefix corresponding to h_1 and $oh'_1 = \mathcal{Z}'_{\mathsf{min}}(h'_1)$. Let $oh_1 = \mathcal{Z}_{\mathsf{min}}(h_1)$ and $s_1 = \mathrm{last}(h'_1)$.

$$\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h')) = \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h'_1)) \times \sum_{a \in A_{\mathsf{min}}} (\Phi_{\mathsf{min}}(\pi)(oh'_1)(s_1)(a) \times \delta'(s_1, a)(s))\,.$$

By the induction hypothesis, $\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h'_1)) = \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h_1))$. For $a \in A_{\mathsf{min}}$: By definition of Φ_{min}, $oh'_1 = (\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh_1)$ we have $\Phi_{\mathsf{min}}(\pi)(oh'_1)(s_1)(a) = \pi(oh_1)(s_1)(a)$ and $\mathrm{last}(h_1) = \mathrm{last}(h'_1) = s_1$ by construction and $\delta' = \delta$ on $S_{\mathsf{min}} \times A_{\mathsf{min}}$ we have $\delta'(s_1, a)(s) = \delta(s_1, a)(s)$. Hence

$$\begin{aligned}
\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h')) &= \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h_1)) \times \sum_{a \in A_{\mathsf{min}}} (\pi(oh_1)(s_1)(a) \times \delta(s_1, a)(s)) \\
&= \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h))
\end{aligned}$$

Case 2: $s \in S_{\mathsf{min}}$. Then $\mathrm{last}(h_1) \in S_{\mathsf{max}}$. By construction, $h' = h'_1, d_s, s$, so $\mathcal{Z}_{\mathsf{max}}(h') = oh'_1, o^\dagger_o, o$ where $o = \mathcal{Z}_{\mathsf{max}}(s)$ and $oh'_1 = \mathcal{Z}'_{\mathsf{max}}(h'_1)$. Let $oh_1 = \mathcal{Z}_{\mathsf{max}}(h_1)$ and $s_1 = \mathrm{last}(h'_1) = \mathrm{last}(h_1)$.

$$\begin{aligned}\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h')) = \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h'_1)) \\ \times \sum_{a\in A_{\mathsf{max}}} (\Phi_{\mathsf{max}}(\sigma)(oh'_1)(s_1)(a) \times \delta'(s_1,a)(d_s)) \\ \times \Phi_{\mathsf{max}}(\sigma)(oh'_1, o^\dagger_o)(d_s)(\top) \times \delta'(d_s,\top)(s).\end{aligned}$$

By the induction hypothesis the $\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h'_1)) = \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h_1))$. For $a \in A_{\mathsf{max}}$: By definition of Φ_{max}, $oh_1 = (\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh'_1)$ $\Phi_{\mathsf{max}}(\sigma)(oh'_1)(s_1)(a) = \sigma(oh_1)(s_1)(a)$ and By construction of $\mathcal{G}'$, $\delta'(s_1,a)(d_s) = \delta(s_1,a)(s)$. Since $\top$ is the sole effective action at any dummy state, $\Phi_{\mathsf{max}}(\sigma)(oh'_1, o^\dagger_o)(d_s)(\top) = 1$. Finally, $\delta'(d_s,\top)(s) = 1$ by construction of $\mathcal{G}'$. Hence

$$\begin{aligned}\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(\mathrm{Cyl}(h')) &= \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h_1)) \times \sum_{a\in A_{\mathsf{max}}} (\sigma(oh_1)(s_1)(a) \times \delta(s_1,a)(s)) \times 1 \times 1 \\ &= \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(\mathrm{Cyl}(h))\end{aligned}$$

The symmetric direction (replacing $\Phi_\star$ by $\Psi_\star = (\Phi_\star)^{-1}$) follows by an identical argument. □

We now state the main correctness theorem for Pre-min transformation.

Lemma B.11. *Let $\mathcal{G}$ be any alternating control POSG and $\mathcal{G}'$ its Pre-*min *transformation. Let $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ be parity objective and $W' = \Gamma(W)$ its corresponding objective in $\mathcal{G}'$. Then*

$$\mathrm{Val}_{\mathcal{G}}(W) = \mathrm{Val}_{\mathcal{G}'}(W').$$

Proof. For fixed $\sigma \in \Sigma_{\mathcal{G}}$, set

$$q_{\mathcal{G}}(\sigma) = \inf_{\pi\in\Pi_{\mathcal{G}}} \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W), \qquad q_{\mathcal{G}'}(\sigma') = \inf_{\pi'\in\Pi_{\mathcal{G}'}} \mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(W').$$

The first equality in Lemma B.10 gives $q_{\mathcal{G}'}(\Phi_{\mathsf{max}}(\sigma)) \le q_{\mathcal{G}}(\sigma)$ because every $\pi \in \Pi_{\mathcal{G}}$ has an image $\Phi_{\mathsf{min}}(\pi) \in \Pi_{\mathcal{G}'}$. Conversely, for every $\pi' \in \Pi_{\mathcal{G}'}$, the second equality there and $\Psi_{\mathsf{max}}(\Phi_{\mathsf{max}}(\sigma)) = \sigma$ give

$$\mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\pi'}_{\mathcal{G}'}(W') = \mathbb{P}^{\sigma,\Psi_{\mathsf{min}}(\pi')}_{\mathcal{G}}(W),$$

so $q_{\mathcal{G}}(\sigma) \le q_{\mathcal{G}'}(\Phi_{\mathsf{max}}(\sigma))$. Hence the two fixed-strategy infima are equal. Since Φ_{max} is a bijection by Lemma B.8, taking the outer supremum yields

$$\begin{aligned}\mathrm{Val}_{\mathcal{G}}(W) &= \sup_{\sigma\in\Sigma_{\mathcal{G}}} q_{\mathcal{G}}(\sigma) = \sup_{\sigma\in\Sigma_{\mathcal{G}}} q_{\mathcal{G}'}(\Phi_{\mathsf{max}}(\sigma)) \\ &= \sup_{\sigma'\in\Sigma_{\mathcal{G}'}} q_{\mathcal{G}'}(\sigma') = \mathrm{Val}_{\mathcal{G}'}(W').\end{aligned}$$

□

Corollary B.12. *Let $\mathcal{G}$ be any alternating control POSG and $\mathcal{G}'$ its Pre-*min *transformation. Let p- a priority function for $\mathcal{G}$. Then*

$$\mathrm{Val}_{\mathcal{G}}\left(\mathrm{Parity}_{\mathcal{G}}(p)\right) = \mathrm{Val}_{\mathcal{G}'}\left(\mathrm{Parity}_{\mathcal{G}'}(p')\right).$$

Proof. Follows directly from lemma B.11 together with Lemma B.9. □

Lemma B.13. *For any $\sigma \in \Sigma_{\mathcal{G}}$ and $\pi \in \Pi_{\mathcal{G}}$:*

$$\Gamma\left(\mathsf{Plays}^{\sigma,\pi}_{\mathcal{G}}\right) = \mathsf{Plays}^{\Phi_{\mathsf{max}}(\sigma),\,\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}.$$

where $\mathsf{Plays}^{\sigma,\pi}_{\mathcal{G}}(resp.\mathsf{Plays}^{\sigma',\pi'}_{\mathcal{G}'})$ denote the set of plays in $\mathcal{G}(resp.\mathcal{G}')$ under the pair of strategies $\sigma,\pi(resp.\sigma',\pi')$.

Proof. Let $\rho' \in \Gamma(\mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi})$. Then there exists $\rho \in \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}$ with $\Gamma(\rho) = \rho'$. Write $\rho = s_0, a_0, s_1, a_1, \ldots$ and $\rho' = s_0, a_0, d_{s_1}, \top, s_1, a_1, \ldots$ as constructed by Γ. We verify ρ' is consistent with $(\Phi_{\mathsf{max}}(\sigma), \Phi_{\mathsf{min}}(\pi))$.

- At each max-**state** $s_k \in S_{\mathsf{max}}$ (k is even): The max-player observation history in $\mathcal{G}'$ at state s_k is $oh'_k = \Gamma_{\mathcal{O}}^{\mathsf{max}}(oh_k)$, where oh is the max-player observation history of ρ at s_k. By definition of Φ_{max}:

$$\begin{aligned}\Phi_{\mathsf{max}}(\sigma)(oh'_k)(s_k) &= \sigma\big((\Gamma_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh'_k)\big)(s_k) \\ &= \sigma(oh_k)(s_k) && \text{from Lemma } B.6\end{aligned}$$

  Because $\rho \in \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}$, action a_k is in the support of $\sigma(oh_k)(s_k)$, and hence also in the support of $\Phi_{\mathsf{max}}(\sigma)(oh'_k)(s_k)$. The max-transition in $\mathcal{G}'$ sends (s_k, a_k) to $d_{s_{k+1}}$ with probability $\delta'(s_k, a_k)(d_{s_{k+1}}) = \delta(s_k, a_k)(s_{k+1})$, matching the transition used in ρ.

- At each **dummy state** $d_{s_{2k+1}} \in D$**:** By definition of Φ_{max}, the strategy assigns $\mathrm{Dirac}(\top)$ to the sole effective action; all other globally enabled actions are max-losing defaults. Moreover, $\delta'(d_{s_{2k+1}}, \top)(s_{2k+1}) = 1$.

- At each min-**state** $s_k \in S_{\mathsf{min}}$ (k is odd): The min-player observation history in $\mathcal{G}'$ at s_k is $oh' = \Gamma_{\mathcal{O}}^{\mathsf{min}}(oh)$, where oh is the min-player observation history of ρ at state s_k. By definition of Φ_{min}:

$$\begin{aligned}\Phi_{\mathsf{min}}(\pi)(oh')(s_k) &= \pi\big((\Gamma_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh')\big)(s_k) \\ &= \pi(oh)(s_k) && \text{from lemma } B.6\end{aligned}$$

  Because $\rho \in \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}$, action a_k is in the support of $\pi(oh)(s_k)$, hence action a_k is also in the support of $\Phi_{\mathsf{min}}(\pi)(oh')(s_k)$.

Thus $\rho' \in \mathsf{Plays}_{\mathcal{G}'}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}$, hence $\Gamma\big(\mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}\big) \subseteq \mathsf{Plays}_{\mathcal{G}'}^{\Phi_{\mathsf{max}}(\sigma),\,\Phi_{\mathsf{min}}(\pi)}$.
Since $\Gamma : \mathsf{Plays}_{\mathcal{G}} \to \mathsf{Plays}_{\mathcal{G}'}$ is a bijection (Lemma $B.5$), every $\rho' \in \mathsf{Plays}_{\mathcal{G}'}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}$ has a unique pre-image $\rho = \Gamma^{-1}(\rho')$ obtained by removing all dummy states and $\top$-actions. By the same step-by-step argument with $(\Gamma_{\mathcal{O}}^{\mathsf{max}})^{-1}$, $(\Gamma_{\mathcal{O}}^{\mathsf{min}})^{-1}$ in place of $\Gamma_{\mathcal{O}}^{\mathsf{max}}$, $\Gamma_{\mathcal{O}}^{\mathsf{min}}$, and using $\Psi_\star = (\Phi_\star)^{-1}$ in place of $\Phi_\star$, the play ρ is consistent with (σ, π), hence $\mathsf{Plays}_{\mathcal{G}'}^{\Phi_{\mathsf{max}}(\sigma),\,\Phi_{\mathsf{min}}(\pi)} \subseteq \Gamma\big(\mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}\big)$. □

Lemma B.14 (Sure Winning)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ be a parity objective of $\mathcal{G}$ and $W' = \Gamma(W)$ be its lifted objective in $\mathcal{G}'$. Let $\sigma \in \Sigma_{\mathcal{G}}$ and $\sigma' = \Phi_{\mathsf{max}}(\sigma)$. Then,*

$$\sigma \textit{ is sure winning for } W \iff \sigma' \textit{ is sure winning for } W'$$

Proof. We will show that

$$\forall \pi \in \Pi_{\mathcal{G}} : \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi} \subseteq W \iff \forall \pi' \in \Pi_{\mathcal{G}'} : \mathsf{Plays}_{\mathcal{G}'}^{\sigma',\pi'} \subseteq W'$$

For forward direction, let $\sigma \in \Sigma_{\mathcal{G}}$ be sure winning for W. Then, $\forall \pi \in \Pi_{\mathcal{G}} : \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi} \subseteq W$. From lemma B.9, for any $\rho \in \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}, \rho \in W \iff \Gamma(\rho) \subseteq W'$. Since ρ is arbitrary, we have $\Gamma(\mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}) \subseteq W'$.
For any $\pi' \in \Pi_{\mathcal{G}'}$: Due to bijection of $\Phi_{\mathsf{min}}, \Psi_{\mathsf{min}}, \exists \pi$ such that $\pi' = \Phi_{\mathsf{min}}(\pi)$. Then, from lemma B.13 and B.8

$$\mathsf{Plays}_{\mathcal{G}'}^{\sigma',\pi'} = \Gamma(\mathsf{Plays}_{\mathcal{G}}^{\Psi_{\mathsf{max}}(\sigma'),\Psi_{\mathsf{min}}(\pi')}) = \Gamma(\mathsf{Plays}_{\mathcal{G}}^{\Psi_{\mathsf{max}}(\Phi_{\mathsf{max}}(\sigma)),\Psi_{\mathsf{min}}(\Phi_{\mathsf{min}}(\pi))}) = \Gamma(\mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}) \subseteq W'$$

The reverse direction can be proven similarly using inverse mappings $\Psi_{\mathsf{max}}, \Psi_{\mathsf{min}}$.

□

Lemma B.15 (Almost-sure Winning)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ be a parity objective of $\mathcal{G}$ and $W' = \Gamma(W)$ be lifted objective in $\mathcal{G}'$. Let $\sigma \in \Sigma_{\mathcal{G}}$ and $\sigma' = \Phi_{\mathsf{max}}(\sigma)$. Then,*

$$\sigma \textit{ is almost-sure winning for } W \iff \sigma' \textit{ is almost-sure winning for } W'$$

Proof. The fixed-strategy equality established in the proof of Lemma B.11 is

$$\inf_{\pi\in\Pi_{\mathcal{G}}} \mathbb{P}_{\mathcal{G}}^{\sigma,\pi}(W) = \inf_{\pi'\in\Pi_{\mathcal{G}'}} \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(W').$$

Hence one side equals 1 if and only if the other does. □

Lemma B.16 (Limit-Sure Winning)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ be a parity objective of $\mathcal{G}$ and $W' = \Gamma(W)$ is the corresponding objective in $\mathcal{G}'$. Then,*

$$W \text{ is limit-sure} \iff W' \text{ is limit-sure}$$

Proof. Follows directly from lemma B.11. □

Lemma B.17 (Quantitative Analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{G}}$ be a parity objective of $\mathcal{G}$ and $W' = \Gamma(W)$ is the corresponding objective in $\mathcal{G}'$. Then, for $k \in (0,1)$*

$$\mathrm{Val}_{\mathcal{G}}(W) \geq k \iff \mathrm{Val}_{\mathcal{G}'}(W') \geq k$$

Proof. Follows directly from lemma B.11. □

B.3.1 Size and Memory Preservation of the reduction

Lemma B.18. *Let $\mathcal{G}$ be an alternating-control POSG and $\mathcal{G}'$ its Pre-*min *transformation. The maps $\Gamma : \mathsf{Plays}_{\mathcal{G}} \to \mathsf{Plays}_{\mathcal{G}'}$, $\Gamma_O^{\mathsf{max}}, \Gamma_O^{\mathsf{min}}$ on observation histories, and $\Phi_\star, \Psi_\star = (\Phi_\star)^{-1}$ ($\star \in \{\mathsf{max}, \mathsf{min}\}$) on strategies satisfy the following. Exact counts use the displayed transitions; the implicit sink completion has constant descriptive overhead. Let $S' = S'_{\mathsf{max}} \cup S'_{\mathsf{min}}$ and $A' = A'_{\mathsf{max}} \cup A'_{\mathsf{min}}$.*

1. (size) $|S'| = |S_{\mathcal{G}}| + |S_{\mathsf{min}}|$, $|A'| = |A_{\mathcal{G}}| + 1$, $|\mathcal{O}'_{\mathsf{max}}| = 2|\mathcal{O}_{\mathsf{max}}|$, $|\mathcal{O}'_{\mathsf{min}}| = 2|\mathcal{O}_{\mathsf{min}}|$, *and* $|\delta'| = |\delta| + |S_{\mathsf{min}}|$. *In particular,* $|\mathcal{G}'| = \mathcal{O}(|\mathcal{G}|)$ *and the construction runs in linear time.*
2. (history length) *For every* $oh \in \mathsf{OH}_\star^{\mathcal{G}}$, $|\Gamma_O^\star(oh)| = \left\lceil \frac{3|oh|-1}{2} \right\rceil$ *and for every* $oh' \in \mathsf{OH}_\star^{\mathcal{G}'}$, $|(\Gamma_O^\star)^{-1}(oh')| \leq |oh'|$. *Observation histories on the two sides are linear in each other.*
3. (memory) *The lifting maps $\Phi_\star, \Psi_\star$ preserve memory class exactly: memoryless, finite-memory and infinite-memory regimes correspond exactly between $\mathcal{G}$ and $\mathcal{G}'$.*

Proof. **Clause (1).** Direct from the construction. The state space gains one fresh dummy state per S_{min}-state, so $|S'| = |S_{\mathsf{max}}| + 2|S_{\mathsf{min}}| = |S_{\mathcal{G}}| + |S_{\mathsf{min}}|$. The action set gains the single fresh action $\top$, so $|A'| = |A_{\mathcal{G}}| + 1$. The observation alphabets are $\mathcal{O}'_{\mathsf{max}} = \mathcal{O}_{\mathsf{max}} \cup \{o_x^\dagger \mid x \in \mathcal{O}_{\mathsf{max}}\}$ and $\mathcal{O}'_{\mathsf{min}} = \mathcal{O}_{\mathsf{min}} \cup \{o_x^\# \mid x \in \mathcal{O}_{\mathsf{min}}\}$, which doubles each cardinality. The transition function δ' agrees with δ on every $(s,a) \in S_{\mathsf{min}} \times A_{\mathsf{min}}$, and on each $(s,a) \in S_{\mathsf{max}} \times A_{\mathsf{max}}$ replaces the target $s_2 \in S_{\mathsf{min}}$ by $d_{s_2} \in D$ – giving the same number of edges as δ – and additionally contributes the deterministic edge $\delta'(d_s, \top) = \mathrm{Dirac}(s)$ for every $d_s \in D$, of which there are $|S_{\mathsf{min}}|$. Hence $|\delta'| = |\delta| + |S_{\mathsf{min}}|$. Each component is therefore a constant or unit-additive perturbation of the corresponding component in $\mathcal{G}$, so $|\mathcal{G}'| = \mathcal{O}(|\mathcal{G}|)$ and $\mathcal{G}'$ is computed from $\mathcal{G}$ in linear time.

Clause (2). Let $\star = \mathsf{max}$ and let $oh = o_0, o_1, \ldots, o_{2n} \in \mathsf{OH}_\star^{\mathcal{G}}$, so that $|oh| = 2n+1$. By construction,

$$\Gamma_O^\star(oh) = o_0, z_1, o_1, o_2, z_3, o_3, \ldots, o_{2n-2}, z_{2n-1}, o_{2n-1}, o_{2n},$$

where z_{2k+1} is the appropriately tagged copy of o_{2k+1} ($z_{2k+1} = o^\dagger_{o_{2k+1}}$) one tagged observation is inserted between every pair of consecutive original observations. Hence

$$|\Gamma_O^\star(oh)| = (2n+1) + n = 3n+1 = \tfrac{3|oh|-1}{2}$$

Let $\star = \mathsf{min}$ and let $oh = o_0, o_1, \ldots, o_{2n+1} \in \mathsf{OH}_\star^{\mathcal{G}}$, so that $|oh| = 2n+2$. By construction,

$$\Gamma_O^\star(oh) = o_0, z_1, o_1, o_2, z_3, o_3, \ldots, o_{2n-2}, z_{2n-1}, o_{2n-1}, o_{2n}, , z_{2n+1}, o_{2n+1}$$

where z_{2k+1} is the appropriately tagged copy of o_{2k+1} ($z_{2k+1} = o^\#_{o_{2k+1}}$) one tagged observation is inserted between every pair of consecutive original observations. Hence

$$|\Gamma_O^\star(oh)| = (2n+2) + n + 1 = 3n+3 = \tfrac{3|oh|}{2}.$$

From above values, for $\star \in \{\mathsf{max}, \mathsf{min}\}$, $|\Gamma^{\star}_{O}(oh)| = \left\lceil \frac{3|oh|-1}{2} \right\rceil$

For the reverse direction, $(\Gamma^{\star}_{O})^{-1}$ acts on $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\star}$ by deleting observations in $\mathcal{O}^{\dagger}$ for $\star = \mathsf{max}$ and in $\mathcal{O}^{\#}$ for $\star = \mathsf{min}$. Deletion only shortens the sequence, so $|(\Gamma^{\star}_{O})^{-1}(oh')| \leq |oh'|$. Combining the two bounds, observation histories on the two sides are within a factor of $\frac{3}{2}$ in length.

Clause (3). Fix $\star \in \{\mathsf{max}, \mathsf{min}\}$ and let a finite-memory strategy be realized by a transducer $M = (Q, q_0, u, g)$, where u updates the memory state from observations and g outputs the complete state-indexed action assignment prescribed by the strategy.

For the forward map $\Phi_{\star}$, run M on $(\Gamma^{\star}_{O})^{-1}(oh')$. This can be done online with the same memory set Q: an inserted observation in $\mathcal{O}^{\dagger}$ or $\mathcal{O}^{\#}$ is skipped, while every original observation is processed exactly as in M. At an original player state, the output is the assignment produced by g; at a dummy max-state, the output assigns $\mathrm{Dirac}(\top)$ to every compatible dummy state. Thus the resulting transducer realizes $\Phi_{\star}$ and uses exactly the memory states in Q.

Conversely, let M' realize a strategy in $\mathcal{G}'$. To realize $\Psi_{\star}$ in $\mathcal{G}$, run M' on the virtual history $\Gamma^{\star}_{O}(oh)$. Each inserted observation is determined by the adjacent original observation, so the corresponding finite block can be fed to M' as soon as that original observation is read; no additional persistent memory is required. The output at an original state is then precisely M''s output after reading $\Gamma^{\star}_{O}(oh)$. Lemmas B.6 and B.7 show that these simulations realize the stated strategy maps.

Therefore both maps preserve the number of memory states. In particular, one-state memoryless strategies, finite-memory strategies, and unrestricted infinite-memory strategies correspond in both directions.

□

B.4 Proof of Lemma 3.2

Proof. Because parity objective subsumes all ω-regular objectives, from lemma B.11, $\mathrm{Val}_{\mathcal{G}}(W) = \mathrm{Val}_{\mathcal{G}'}(W')$ for any ω-regular objective W in $\mathcal{G}$ and its corresponding objective W' in $\mathcal{G}'$.
The "consequently" statement follows from lemmas B.14, B.15, B.16, and B.17. □

C Constructing an RPOMDP from a Pre-min Transformed POSG

Let $\mathcal{G}'$ be the Pre-min transformation of an alternating control POSG $\mathcal{G}$ as constructed in definition B.1. In $\mathcal{G}'$, every play has the periodic structure

$$s_1 \xrightarrow{a_1} d_{s_2} \xrightarrow{\top} s_2 \xrightarrow{a_2} s_3 \xrightarrow{a_3} d_{s_4} \xrightarrow{\top} s_4 \cdots$$

where max-states and dummy states in $S_{\mathsf{max}} \cup D$ alternate with S_{min}-states. The key observation is that the min-player's choice of action $a \in A_{\mathsf{min}}$ at a S_{min}-state s determines the next transition distribution $\delta'(s,a) \in \Delta(S_{\mathsf{max}})$. This is exactly the role of the *environment* in an RPOMDP: given a state and an action, it selects a distribution from the uncertainty set. We exploit this to define a polytopic RPOMDP $\mathfrak{R}(\mathcal{G})$ whose states are the max-states and dummy states of $\mathcal{G}'$, and whose uncertainty set at each dummy state under action $\top$ is the *polytope* whose vertices are the transition distributions corresponding to the min-player's available actions.

Definition C.1 (RPOMDP reduction of a Pre-min transformed POSG)**.** Let $\mathcal{G}' = (S_{\mathsf{max}} \cup D, S_{\mathsf{min}}, A_{\mathsf{max}} \cup \{\top\}, A_{\mathsf{min}}, \delta', s_0, \mathcal{O}_{\mathsf{max}} \cup \mathcal{O}^{\dagger}, \mathcal{O}_{\mathsf{min}} \cup \mathcal{O}^{\#}, \mathcal{Z}'_{\mathsf{max}}, \mathcal{Z}'_{\mathsf{min}})$ be a Pre-min transformed POSG. We define the associated RPOMDP as

$$\mathfrak{R}(\mathcal{G}) = (S_{\mathfrak{R}},\ A_{\mathfrak{R}},\ \mathcal{P}_{\mathfrak{R}},\ \mathcal{O}^{\mathfrak{R}}_{\mathsf{a}},\ \mathcal{O}^{\mathfrak{R}}_{\mathsf{e}},\ \mathcal{Z}^{\mathfrak{R}}_{\mathsf{a}},\ \mathcal{Z}^{\mathfrak{R}}_{\mathsf{e}},\ s^{\mathfrak{R}}_0),$$

where:

- *(State space)* $S_{\mathfrak{R}} = S_{\mathsf{max}} \cup D$.
- *(Actions)* $A_{\mathfrak{R}} = A_{\mathsf{max}} \cup \{\top\}$, with every action enabled at every state. The effective pairs are $(s,a) \in S_{\mathsf{max}} \times A_{\mathsf{max}}$ and $(d_s, \top)$; all other pairs use the default max-losing singleton transition.
- *(Uncertainty set)* $\mathcal{P}_{\mathfrak{R}} = \prod_{(s,a) \in S_{\mathfrak{R}} \times A_{\mathfrak{R}}} P(s,a)$, where the non-default polytopes are:
  - For $(s,a) \in S_{\mathsf{max}} \times A_{\mathsf{max}}$ or $s = d_{s'} \in D$ and $a = \top$, define

$$V_{s,a} = \begin{cases} \{\delta'(s,a)\} & \text{if } s \in S_{\mathsf{max}},\ a \in A_{\mathsf{max}}, \\ \{\, \delta'(s',a') \mid a' \in A_{\mathsf{min}} \,\} & \text{if } s = d_{s'},\ a = \top,\ s' \in S_{\mathsf{min}}. \end{cases}$$

    For these pairs,

$$P(s,a) = \mathrm{conv}(V_{s,a}),$$

    where $\mathrm{conv}(V_{s,a})$ denotes the convex hull of $V_{s,a}$. Every remaining pair uses the implicit max-losing singleton.
- *(Agent observations)* $\mathcal{O}^{\mathfrak{R}}_{\mathsf{a}} = \mathcal{O}_{\mathsf{max}} \cup \mathcal{O}^{\dagger}$.
- *(Environment observations)* $\mathcal{O}^{\mathfrak{R}}_{\mathsf{e}} = \mathcal{O}_{\mathsf{min}} \cup \mathcal{O}^{\#}$.
- *(Agent observation function)* $\mathcal{Z}^{\mathfrak{R}}_{\mathsf{a}} : S_{\mathfrak{R}} \to \mathcal{O}^{\mathfrak{R}}_{\mathsf{a}}$ with $\mathcal{Z}^{\mathfrak{R}}_{\mathsf{a}}(s') = \mathcal{Z}'_{\mathsf{max}}(s')$ for $s' \in S_{\mathfrak{R}}$.
- *(Environment observation function)* $\mathcal{Z}^{\mathfrak{R}}_{\mathsf{e}} : S_{\mathfrak{R}} \to \mathcal{O}^{\mathfrak{R}}_{\mathsf{e}}$ with $\mathcal{Z}^{\mathfrak{R}}_{\mathsf{e}}(s') = \mathcal{Z}'_{\mathsf{min}}(s')$ for $s' \in S_{\mathfrak{R}}$.
- *(Initial state)* $s^{\mathfrak{R}}_0 = s_0$.

Remark C.2. *The vertex set $V_{d_s,\top} = \{\delta'(s,a) \mid a \in A_{\mathsf{min}}\}$ of the polytope $P(d_s,\top)$ encodes all* min*-actions at $s \in S_{\mathsf{min}}$: each vertex corresponds to the distribution induced by a* min*-action. The environment in $\mathfrak{R}(\mathcal{G})$ selects any convex combination of these vertices, which corresponds to the* min*-player in $\mathcal{G}'$ playing a randomized distribution over actions at state s. At non-dummy states $s' \in S_{\mathsf{max}}$, the uncertainty set is a singleton, so the environment has no real choice there. This makes $\mathfrak{R}(\mathcal{G})$ a polytopic (s,a)-rectangular RPOMDP.*

C.1 Mapping plays, observation histories, strategies from $\mathcal{G}'$ to $\mathfrak{R}(\mathcal{G})$

Mapping on plays. Because the uncertainty set at dummy states is polytopic (not a singleton), different plays in $\mathcal{G}'$ that agree on all $S_{\mathsf{max}} \cup D$ states and $A_{\mathsf{max}} \cup \{\top\}$ actions but differ in the min-action chosen at S_{min}-states will map to the *same* play in $\mathfrak{R}(\mathcal{G})$. We therefore define $\Lambda : \mathsf{Plays}_{\mathcal{G}'} \to \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}$ and its set-valued inverse $\Lambda^{-1} : \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})} \to 2^{\mathsf{Plays}_{\mathcal{G}'}}$.

Let $\rho' \in \mathsf{Plays}_{\mathcal{G}'}$ be of the form $\rho' = s_1, a_1, s_2, a_2, s_3, a_3, \ldots$ where by construction of $\mathcal{G}'$, for all $k \geq 0$:

$$s_{3k+1} \in S_{\mathsf{max}}, \quad s_{3k+2} \in D, \quad s_{3k+3} \in S_{\mathsf{min}}, \quad a_{3k+1} \in A_{\mathsf{max}}, \quad a_{3k+2} = \top, \quad a_{3k+3} \in A_{\mathsf{min}}.$$

We define $\Lambda(\rho') = \rho^{\mathfrak{R}}$ where $\rho^{\mathfrak{R}} = t_1, b_1, t_2, b_2, \ldots \in \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}$ is given index-wise, for all $k \geq 0$, by:

$$\begin{aligned} t_{2k+1} &= s_{3k+1} \in S_{\mathsf{max}}, \\ b_{2k+1} &= a_{3k+1} \in A_{\mathsf{max}}, \\ t_{2k+2} &= s_{3k+2} \in D, \\ b_{2k+2} &= \top. \end{aligned}$$

That is, we remove all occurrences of S_{min}-states and A_{min}-actions from ρ'.

Set-valued inverse. Define $\Lambda^{-1} : \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})} \to 2^{\mathsf{Plays}_{\mathcal{G}'}}$ as follows. Let $\rho^{\mathfrak{R}} = t_1, b_1, t_2, b_2, \ldots \in \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}$ where for all $k \geq 0$: $t_{2k+1} \in S_{\mathsf{max}}$, $t_{2k+2} \in D$, $b_{2k+1} \in A_{\mathsf{max}}$, $b_{2k+2} = \top$. Then $\Lambda^{-1}(\rho^{\mathfrak{R}})$ is the set of all plays $\rho' = s_1, a_1, s_2, a_2, s_3, a_3, \ldots \in \mathsf{Plays}_{\mathcal{G}'}$ such that for all $k \geq 0$:

$$\begin{aligned} &s_{3k+1} = t_{2k+1}, \\ &a_{3k+1} = b_{2k+1}, \\ &s_{3k+2} = t_{2k+2}, \\ &a_{3k+2} = \top, \\ &s_{3k+3} = s \text{ such that } t_{2k+2} = d_s \in D, \\ &a_{3k+3} \in A_{\mathsf{min}} \text{ (arbitrary action subject to } \delta'(s_{3k+3}, a_{3k+3})(s_{3k+4}) > 0). \end{aligned}$$

Thus $\Lambda^{-1}(\rho^{\mathfrak{R}})$ consists of all plays in $\mathcal{G}'$ obtained by resolving each transition at a dummy state using an arbitrary min-action. In particular, $\Lambda(\rho') = \rho^{\mathfrak{R}}$ for every $\rho' \in \Lambda^{-1}(\rho^{\mathfrak{R}})$, i.e. Λ is many-to-one.

Mapping on observation histories. We extend Λ to observation histories exactly as before. Let $oh = o_1, o_2, o_3, \ldots, o_n \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$.

We define $\Lambda^{\mathsf{max}}_{\mathcal{O}} : \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}} \to \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}}$ by $\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh) = oh^{\mathfrak{R}}$ where $oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \tilde{o}_3, \ldots$ is given index-wise, for all $k \geq 0$ such that indices exist, by:

$$\begin{aligned} \tilde{o}_{2k+1} &= o_{3k+1}, \\ \tilde{o}_{2k+2} &= o_{3k+2}. \end{aligned}$$

Observations o_{3k+3} of S_{min}-states are removed. The map $\Lambda^{\mathsf{max}}_{\mathcal{O}}$ is a bijection with inverse $(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1} : \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}} \to \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$ defined, for $oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_m \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}}$ ($\tilde{o}_{2k+1} \in \mathcal{O}_{\mathsf{max}}, \tilde{o}_{2k+2} \in o^{\dagger}$ for $k \geq 0$), by $(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}}) = oh = o_1, o_2, \ldots, o_k$ where for all $k \geq 0$:

$$\begin{aligned} o_{3k+1} &= \tilde{o}_{2k+1}, \\ o_{3k+2} &= \tilde{o}_{2k+2}, \\ o_{3k+3} &= x \text{ such that } \tilde{o}_{2k+2} = o^{\dagger}_x, \\ o_k &= \tilde{o}_m \end{aligned}$$

We extend this to the min-player. We define $\Lambda^{\mathsf{min}}_{\mathcal{O}} : \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}} \to \widehat{\mathsf{OH}}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}}$, where $\widehat{\mathsf{OH}}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}} \subseteq \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}}$ is the set of environment observation histories of $\mathfrak{R}(\mathcal{G})$ ending in an observation of a state in D, i.e, $\widehat{\mathsf{OH}}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}} = \{oh^{\mathfrak{R}} \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}} \mid \mathrm{last}(oh^{\mathfrak{R}}) \in \mathcal{O}^{\#}\}$. For $oh = o_1, o_2, \ldots, o_n \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}}$, $\Lambda^{\mathsf{min}}_{\mathcal{O}}(oh) = \tilde{o}_1, \tilde{o}_2, \ldots$ is defined index-wise for all $k \geq 0$ by:

$$\begin{aligned} \tilde{o}_{2k+1} &= o_{3k+1}, \\ \tilde{o}_{2k+2} &= o_{3k+2}. \end{aligned}$$

Since $\mathrm{last}(oh) \in \mathcal{O}_{\mathsf{min}}$, the mapped history ends in $\mathcal{O}^{\#}$ and hence lies in $\widehat{\mathsf{OH}}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}}$. Its inverse $(\Lambda^{\mathsf{min}}_{\mathcal{O}})^{-1} : \widehat{\mathsf{OH}}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}} \to \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}}$ is defined for $oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_m \in \widehat{\mathsf{OH}}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{e}}$ by $(\Lambda^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}}) =$

$o_1, o_2, \ldots, o_{m+m/2}$ where for all $k \geq 0$:

$$\begin{aligned} o_{3k+1} &= \tilde{o}_{2k+1}, \\ o_{3k+2} &= \tilde{o}_{2k+2}, \\ o_{3k+3} &= x \text{ such that } \tilde{o}_{2k+2} = o_x^{\#}, \\ o_{m+m/2} &= x \text{ such that } \tilde{o}_m = o_x^{\#}. \end{aligned}$$

Lemma C.3. *For all* $oh^{\mathfrak{R}} \in \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})}$, $\Lambda_{\mathcal{O}}^{\mathsf{max}}\big((\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh^{\mathfrak{R}})\big) = oh^{\mathfrak{R}}$*; and for all* $oh' \in \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'}$, $(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(\Lambda_{\mathcal{O}}^{\mathsf{max}}(oh')) = oh'$.

Lemma C.4. *For all* $oh^{\mathfrak{R}} \in \widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}$, $\Lambda_{\mathcal{O}}^{\mathsf{min}}\big((\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh^{\mathfrak{R}})\big) = oh^{\mathfrak{R}}$*; and for all* $oh' \in \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}'}$, $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}\big(\Lambda_{\mathcal{O}}^{\mathsf{min}}(oh')\big) = oh'$.

Mapping of strategies. Recall from section 2 and the definition of RPOMDP that:

- A player strategy in $\mathcal{G}'$ maps its observation history to state-indexed distributions over its global action alphabet, as in the main text.
- An agent strategy $\sigma^{\mathfrak{R}} \in \Sigma_{\mathfrak{R}(\mathcal{G})}$ maps each observation history to state-indexed distributions over $A_{\mathfrak{R}}$.
- An environment strategy $\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}(\mathcal{G})}$ has type $\pi^{\mathfrak{R}} : \mathsf{OH}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})} \times A_{\mathfrak{R}} \to \mathcal{P}_{\mathfrak{R}}$ and hence selects a complete transition assignment.

Since $\mathcal{P}_{\mathfrak{R}}$ is (s,a)-rectangular and polytopic, the environment strategy $\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}(\mathcal{G})}$ at a dummy state $d_s \in D$ under action $\top$ picks a distribution $\pi^{\mathfrak{R}}(oh^{\mathfrak{R}}, \top)(d_s, \top) \in P(d_s, \top)$, which by Definition C.1 is a convex combination of the vertices $V_{d_s,\top} = \{\delta'(s,a) \mid a \in A_{\mathsf{min}}\}$. We write this as

$$\pi^{\mathfrak{R}}(oh^{\mathfrak{R}}, \top)(d_s, \top) = \sum_{a \in A_{\mathsf{min}}} \alpha_a \cdot \delta'(s,a), \quad \alpha_a \geq 0, \quad \sum_{a \in A_{\mathsf{min}}} \alpha_a = 1.$$

For every $s \in S_{\mathsf{min}}$, define

$$\mathsf{bar}_s^{\mathcal{G}'} : \Delta(A_{\mathsf{min}}) \to P(d_s, \top), \qquad \mathsf{bar}_s^{\mathcal{G}'}(\lambda) = \sum_{a \in A_{\mathsf{min}}} \lambda(a)\delta'(s,a).$$

For each $P \in P(d_s, \top)$, choose $\beta_s^P \in \Delta(A_{\mathsf{min}})$ such that $\mathsf{bar}_s^{\mathcal{G}'}(\beta_s^P) = P$. The coefficients may be nonunique; fix one such β_s^P for every P. For a complete assignment $P \in \mathcal{P}_{\mathfrak{R}}$, fix a reference element $P_{x,b}^0 \in P(x,b)$ for every $(x,b) \in S_{\mathfrak{R}} \times A_{\mathfrak{R}}$. We define the following four maps lifting strategies between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$:

1. $\Omega_{\mathsf{max}} : \Sigma_{\mathcal{G}'} \to \Sigma_{\mathfrak{R}(\mathcal{G})}$: for each $\sigma' \in \Sigma_{\mathcal{G}'}$, $oh^{\mathfrak{R}} \in \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})}$ and $\forall s \in S_{\mathfrak{R}} = S_{\mathsf{max}} \cup D$ such that $\mathcal{Z}_{\mathsf{a}}^{\mathfrak{R}}(s) = \mathrm{last}(oh^{\mathfrak{R}})$,

$$\Omega_{\mathsf{max}}(\sigma')(oh^{\mathfrak{R}})(s) = \sigma'\big((\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh^{\mathfrak{R}})\big)(s)$$

2. $\Omega_{\mathsf{min}} : \Pi_{\mathcal{G}'} \to \Pi_{\mathfrak{R}(\mathcal{G})}$: for each $\pi' \in \Pi_{\mathcal{G}'}$, $oh^{\mathfrak{R}} \in \mathsf{OH}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}$, and $b \in A_{\mathfrak{R}}$, define the complete assignment $\Omega_{\mathsf{min}}(\pi')(oh^{\mathfrak{R}}, b) \in \mathcal{P}_{\mathfrak{R}}$ by

$$\Omega_{\mathsf{min}}(\pi')(oh^{\mathfrak{R}}, b)(x,c) = \begin{cases} \mathsf{bar}_s^{\mathcal{G}'}\big(\pi'\big((\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh^{\mathfrak{R}})\big)(s)\big) & \text{if } x = d_s,\ b = c = \top, \\ & \mathcal{Z}_{\mathsf{e}}^{\mathfrak{R}}(x) = \mathrm{last}(oh^{\mathfrak{R}}), \\ \delta'(x,b) & \text{if } x \in S_{\mathsf{max}},\ b = c \in A_{\mathsf{max}}, \\ & \mathcal{Z}_{\mathsf{e}}^{\mathfrak{R}}(x) = \mathrm{last}(oh^{\mathfrak{R}}), \\ P_{x,c}^0 & \text{otherwise.} \end{cases}$$

3. $\Xi_{\mathsf{max}} : \Sigma_{\mathfrak{R}(\mathcal{G})} \to \Sigma_{\mathcal{G}'}$: for each $\sigma^{\mathfrak{R}} \in \Sigma_{\mathfrak{R}(\mathcal{G})}$, $oh' \in \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'}$ and $\forall s \in S_{\mathsf{max}} \cup D$ such that $\mathcal{Z}'_{\mathsf{max}}(s) = \mathrm{last}(oh')$,

$$\Xi_{\mathsf{max}}(\sigma^{\mathfrak{R}})(oh')(s) = \sigma^{\mathfrak{R}}(\Lambda_{\mathcal{O}}^{\mathsf{max}}(oh'))(s).$$

4. $\Xi_{\mathsf{min}} : \Pi_{\mathfrak{R}(\mathcal{G})} \to \Pi_{\mathcal{G}'}$: for each $\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}(\mathcal{G})}$, $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}}$ and $\forall s' \in S_{\mathsf{min}}$ with $\mathcal{Z}'_{\mathsf{min}}(s') = \mathrm{last}(oh')$, define

$$\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})(oh')(s') = \beta_{s'}^{\pi^{\mathfrak{R}}(\Lambda^{\mathsf{min}}_{\mathcal{O}}(oh'),\top)(d_{s'},\top)}.$$

Remark C.5. *By definition of Ξ_{min}, for all $\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}(\mathcal{G})}$, $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{min}}$, and $d_{s'} \in D$,*

$$\pi^{\mathfrak{R}}(\Lambda^{\mathsf{min}}_{\mathcal{O}}(oh'),\top)(d_{s'},\top) = \mathsf{bar}^{\mathcal{G}'}_{s'}\left(\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})(oh')(s')\right).$$

Lemma C.6. *The agent maps Ω_{max} and Ξ_{max} are mutual inverses. For every compatible dummy history and state, the environment maps satisfy*

$$\Omega_{\mathsf{min}}(\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}}))(oh^{\mathfrak{R}},\top)(d_s,\top) = \pi^{\mathfrak{R}}(oh^{\mathfrak{R}},\top)(d_s,\top),$$
$$\mathsf{bar}^{\mathcal{G}'}_{s}(\Xi_{\mathsf{min}}(\Omega_{\mathsf{min}}(\pi'))(oh')(s)) = \mathsf{bar}^{\mathcal{G}'}_{s}(\pi'(oh')(s)).$$

Thus the environment compositions preserve transition distributions.

Proof. The agent identities follow immediately from Lemmas C.3 and the definitions of $\Omega_{\mathsf{max}}, \Xi_{\mathsf{max}}$. The environment identities follow from $\mathsf{bar}^{\mathcal{G}'}_{s}(\beta^{P}_{s}) = P$ and Lemma C.4. At non-dummy states the uncertainty set is a singleton. □

C.2 Lifting objectives from $\mathcal{G}'$ to $\mathfrak{R}(\mathcal{G})$

Define $p^{\mathfrak{R}} : S_{\mathfrak{R}} \to \{0, 1, \ldots, d\}, d \in \mathbb{N}$ by $p^{\mathfrak{R}}(s') = p'(s')$ for $s' \in S_{\mathfrak{R}}$ where p' is priority function for $\mathcal{G}'$.

Lemma C.7. *Let $\mathcal{G}'$, $\mathfrak{R}(\mathcal{G})$, p', and $p^{\mathfrak{R}}$ be as above, and let $\rho' \in \mathsf{Plays}_{\mathcal{G}'}$. Then*

$$\rho' \in \mathrm{Parity}_{\mathcal{G}'}(p') \iff \Lambda(\rho') \in \mathrm{Parity}_{\mathfrak{R}(\mathcal{G})}\left(p^{\mathfrak{R}}\right).$$

Proof. By construction of $p^{\mathfrak{R}}$, every state $s' \in S_{\mathsf{max}}$ in $\Lambda(\rho')$ carries priority $p^{\mathfrak{R}}(s') = p'(s')$, and every dummy state $d_s \in D$ carries priority $p^{\mathfrak{R}}(d_s) = p'(d_s)$. The S_{min}-states of ρ' are removed by Λ; since by definition of p' the priority of each $s \in S_{\mathsf{min}}$ equals that of the preceding dummy state d_s, no $\liminf$-relevant priority value is lost. Hence the parity condition is preserved. □

C.3 Value Preservation between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$

Lemma C.8. *The strategy maps preserve objective probabilities in both directions:*

$$\mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(W') = \mathbb{P}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}),$$
$$\mathbb{P}^{\sigma^{\mathfrak{R}},\pi^{\mathfrak{R}}}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}) = \mathbb{P}^{\Xi_{\mathsf{max}}(\sigma^{\mathfrak{R}}),\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})}_{\mathcal{G}'}(W'),$$

for all strategies of the indicated types, where $W' \subseteq \mathsf{Plays}_{\mathcal{G}'}$ is parity objective in $\mathcal{G}'$ and $W^{\mathfrak{R}} = \Lambda(W')$ is the corresponding objective in $\mathfrak{R}(\mathcal{G})$.

Proof. Since the probability measure over plays is uniquely determined by its values on cylinder sets, it suffices to show that for every finite history h' in W' ending at a state in $S_{\mathsf{max}} \cup D$, there exists a corresponding history $h^{\mathfrak{R}}$ in $W^{\mathfrak{R}}$ such that

$$\mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(\mathrm{Cyl}(h')) = \mathbb{P}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}(\mathrm{Cyl}\left(h^{\mathfrak{R}}\right)).$$

We map each such h' to $h^{\mathfrak{R}}$ similar to Λ maps the corresponding plays (dropping S_{min}-states). The proof proceeds by induction on the length of h'.

Base case. $h' = s_0$. Then $h^{\mathfrak{R}} = s_0^{\mathfrak{R}} = s_0$ and both cylinders have probability 1.

Inductive step. We distinguish three cases based on how h' is extended.

Case 1: h' is of the form h'_1, d_s where $d_s \in D$. Then $\mathrm{last}(h'_1) \in S_{\mathsf{max}}$ and $h^{\mathfrak{R}} = h_1^{\mathfrak{R}}, d_s$ where $h_1^{\mathfrak{R}}$ corresponds to h'_1 by induction. Let $oh_a^{\mathfrak{R}} = \mathcal{Z}_{\mathsf{a}}^{\mathfrak{R}}(h_1^{\mathfrak{R}})$. Since $\mathrm{last}(h_1^{\mathfrak{R}}) = s_1 \in S_{\mathsf{max}}$, the uncertainty set for $a \in A_{\mathsf{max}}$, $P(s_1, a)$ is a singleton $\{\delta'(s_1, a)\}$. Then

$$\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h^{\mathfrak{R}})) = \mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h_1^{\mathfrak{R}})) \times \sum_{a \in A_{\mathsf{max}}} \left(\Omega_{\mathsf{max}}(\sigma')(oh_a^{\mathfrak{R}})(s_1)(a) \times \delta'(s_1, a)(d_s)\right).$$

By the induction hypothesis, $\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h_1^{\mathfrak{R}})) = \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'_1))$. Since $\mathrm{last}(oh_a^{\mathfrak{R}}) \in \mathcal{O}_{\mathsf{max}}$ and $(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh_a^{\mathfrak{R}}) = \mathcal{Z}'_{\mathsf{max}}(h'_1)$, we have $\Omega_{\mathsf{max}}(\sigma')(oh_a^{\mathfrak{R}})(s_1)(a) = \sigma'(\mathcal{Z}'_{\mathsf{max}}(h'_1))(s_1)(a)$. The singleton uncertainty set means the environment contributes the unique element $\delta'(s_1, a)$ evaluated for transition to d_s. Hence

$$\begin{aligned}\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h^{\mathfrak{R}})) &= \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'_1)) \times \sum_{a \in A_{\mathsf{max}}} (\sigma'(\mathcal{Z}'_{\mathsf{max}}(h'_1))(s_1)(a) \times \delta'(s_1, a)(d_s)) \\ &= \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'))\end{aligned}$$

Case 2: h' is of the form h'_1, s where $s \in S_{\mathsf{min}}$ and $\mathrm{last}(h'_1) = d_s \in D$. The sole effective action from d_s is $\top$ and the transition $\delta'(d_s, \top) = \mathrm{Dirac}(s)$ is deterministic. The corresponding prefix in $\mathfrak{R}(\mathcal{G})$ is $h^{\mathfrak{R}} = h_1^{\mathfrak{R}}$ (unchanged, since the S_{min}-state s is removed by Λ). Since the $\top$ step at d_s is deterministic (probability 1) in both $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$, and the S_{min}-state s is not present in $\mathfrak{R}(\mathcal{G})$, we have immediately

$$\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h^{\mathfrak{R}})) = \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'))$$

Case 3: h' is of the form h'_1, s, s' where $\mathrm{last}(h'_1) = d_s \in D$, $s \in S_{\mathsf{min}}$, and $s' \in S_{\mathsf{max}}$. Then $h^{\mathfrak{R}} = h_1^{\mathfrak{R}}, s'$ where $h_1^{\mathfrak{R}}$ corresponds to h'_1. Let $oh_a^{\mathfrak{R}} = \mathcal{Z}_{\mathsf{a}}^{\mathfrak{R}}(h_1^{\mathfrak{R}})$ and $oh_e^{\mathfrak{R}} = \mathcal{Z}_{\mathsf{e}}^{\mathfrak{R}}(h_1^{\mathfrak{R}})$. Since $\mathrm{last}(h_1^{\mathfrak{R}}) = d_s \in D$, we have $\mathrm{last}(oh_e^{\mathfrak{R}}) \in \mathcal{O}^{\#}$. The environment in $\mathfrak{R}(\mathcal{G})$ selects a distribution from $P(d_s, \top)$; the mapped agent plays the sole effective action $\top$ with probability 1.

$$\begin{aligned}\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h^{\mathfrak{R}})) = \mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h_1^{\mathfrak{R}})) \times \Omega_{\mathsf{max}}(\sigma')(oh_a^{\mathfrak{R}})(d_s)(\top) \\ \times \Omega_{\mathsf{min}}(\pi')(oh_e^{\mathfrak{R}}, \top)(d_s, \top)(s')\end{aligned}$$

where we use the fact that the distribution selected by the environment at $(d_s, \top)$ is evaluated at s'. By definition of Ω_{min},

$$\Omega_{\mathsf{min}}(\pi')(oh_e^{\mathfrak{R}}, \top)(d_s, \top) = \sum_{a'' \in A_{\mathsf{min}}} \pi'\left((\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh_e^{\mathfrak{R}})\right)(s)(a'') \cdot \delta'(s, a'').$$

Evaluating at s':

$$\Omega_{\mathsf{min}}(\pi')(oh_e^{\mathfrak{R}}, \top)(d_s, \top)(s') = \sum_{a'' \in A_{\mathsf{min}}} \pi'\left((\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh_e^{\mathfrak{R}})\right)(s)(a'') \cdot \delta'(s, a'')(s').$$

Since $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh_e^{\mathfrak{R}}) = \mathcal{Z}'_{\mathsf{min}}(h'_1, s)$, this equals $\sum_{a'' \in A_{\mathsf{min}}} \pi'(\mathcal{Z}'_{\mathsf{min}}(h'_1, s))(s)(a'') \cdot \delta'(s, a'')(s')$. Also, $\Omega_{\mathsf{max}}(\sigma')(oh_a^{\mathfrak{R}})(d_s)(\top) = 1$ since $\top$ is the sole effective action at dummy states. By the induction hypothesis applied to h'_1,

$$\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h^{\mathfrak{R}})) = \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'_1)) \times \sum_{a'' \in A_{\mathsf{min}}} \pi'(\mathcal{Z}'_{\mathsf{min}}(h'_1, s))(s)(a'') \cdot \delta'(s, a'')(s').$$

On the other hand, in $\mathcal{G}'$, from history h'_1 ending at d_s, the max-player plays $\top$ (probability 1), reaching s deterministically, after which the min-player at s chooses action $a'' \in A_{\mathsf{min}}$ with probability $\pi'(\mathcal{Z}'_{\mathsf{min}}(h'_1))(s)(a'')$, reaching s' with probability $\delta'(s, a'')(s')$. Therefore, summing over all min actions in $\mathcal{G}'$:

$$\mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h')) = \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'_1)) \times 1 \times \sum_{a'' \in A_{\mathsf{min}}} \pi'(\mathcal{Z}'_{\mathsf{min}}(h'_1, s))(s)(a'') \cdot \delta'(s, a'')(s').$$

These two expressions agree, so $\mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(\mathrm{Cyl}(h^{\mathfrak{R}})) = \mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(\mathrm{Cyl}(h'))$.

For the converse mapping, repeat the cylinder induction with Ξ_{max} and Ξ_{min}. At a dummy state, the first environment identity in Lemma C.6 ensures that the selected action mixture has barycenter exactly $\pi^{\mathfrak{R}}(oh^{\mathfrak{R}}, \top)(d_s, \top)$. Thus every cylinder probability, and consequently the probability of the parity objective, is preserved in the reverse direction as well. □

We now state the main equivalence theorem for the RPOMDP reduction.

Lemma C.9. *Let $\mathcal{G}'$ be a Pre-*min *transformed POSG and $\mathfrak{R}(\mathcal{G})$ its RPOMDP reduction. Let $W' \subseteq \mathsf{Plays}_{\mathcal{G}'}$ be parity objective and $W^{\mathfrak{R}} = \Lambda(W')$ its corresponding objective in $\mathfrak{R}(\mathcal{G})$. Then*

$$\mathrm{Val}_{\mathcal{G}'}(W') = \mathrm{Val}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}).$$

Proof. For fixed $\sigma' \in \Sigma_{\mathcal{G}'}$, write

$$q_{\mathcal{G}'}(\sigma') = \inf_{\pi' \in \Pi_{\mathcal{G}'}} \mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(W'), \qquad q_{\mathfrak{R}(\mathcal{G})}(\sigma^{\mathfrak{R}}) = \inf_{\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}(\mathcal{G})}} \mathbb{P}^{\sigma^{\mathfrak{R}},\pi^{\mathfrak{R}}}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}).$$

The first equality in Lemma C.8 gives $q_{\mathfrak{R}(\mathcal{G})}(\Omega_{\mathsf{max}}(\sigma')) \leq q_{\mathcal{G}'}(\sigma')$. Conversely, for every $\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}(\mathcal{G})}$, the second equality there and $\Xi_{\mathsf{max}}(\Omega_{\mathsf{max}}(\sigma')) = \sigma'$ give

$$\mathbb{P}^{\Omega_{\mathsf{max}}(\sigma'),\pi^{\mathfrak{R}}}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}) = \mathbb{P}^{\sigma',\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})}_{\mathcal{G}'}(W'),$$

so $q_{\mathcal{G}'}(\sigma') \leq q_{\mathfrak{R}(\mathcal{G})}(\Omega_{\mathsf{max}}(\sigma'))$. Hence the two fixed-strategy infima are equal. Since Ω_{max} is a bijection by Lemma C.6,

$$\begin{aligned}\mathrm{Val}_{\mathcal{G}'}(W') &= \sup_{\sigma' \in \Sigma_{\mathcal{G}'}} q_{\mathcal{G}'}(\sigma') = \sup_{\sigma' \in \Sigma_{\mathcal{G}'}} q_{\mathfrak{R}(\mathcal{G})}(\Omega_{\mathsf{max}}(\sigma')) \\ &= \sup_{\sigma^{\mathfrak{R}} \in \Sigma_{\mathfrak{R}(\mathcal{G})}} q_{\mathfrak{R}(\mathcal{G})}(\sigma^{\mathfrak{R}}) = \mathrm{Val}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}).\end{aligned}$$

□

Corollary C.10. *Let $\mathcal{G}'$ be a Pre-*min *transformed POSG and $\mathfrak{R}(\mathcal{G})$ its RPOMDP reduction. Let p' a priority function for $\mathcal{G}'$. Then*

$$\mathrm{Val}_{\mathcal{G}'}\left(\mathrm{Parity}_{\mathcal{G}'}(p')\right) = \mathrm{Val}_{\mathfrak{R}(\mathcal{G})}\left(\mathrm{Parity}_{\mathfrak{R}(\mathcal{G})}(p^{\mathfrak{R}})\right).$$

Proof. Follows directly from lemma C.9 together with lemma C.7. □

Lemma C.11. *Projected supported plays are preserved in both mapping directions:*

$$\begin{aligned}\Lambda\left(\mathsf{Plays}^{\sigma',\pi'}_{\mathcal{G}'}\right) &= \mathsf{Plays}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}, \\ \Lambda\left(\mathsf{Plays}^{\Xi_{\mathsf{max}}(\sigma^{\mathfrak{R}}),\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})}_{\mathcal{G}'}\right) &= \mathsf{Plays}^{\sigma^{\mathfrak{R}},\pi^{\mathfrak{R}}}_{\mathfrak{R}(\mathcal{G})}.\end{aligned}$$

Proof. We prove both equalities by induction on corresponding finite prefixes. The initial prefixes consist of s_0 on both sides. Suppose first that corresponding prefixes under (σ', π') end at an original max-state $s \in S_{\mathsf{max}}$. Their agent observation histories correspond under $\Lambda^{\mathsf{max}}_{\mathcal{O}}$, and the definition of Ω_{max} gives

$$\Omega_{\mathsf{max}}(\sigma')(\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh'))(s) = \sigma'(oh')(s).$$

Thus the same max-actions have positive probability. For every such action a, the RPOMDP uncertainty set at (s, a) is the singleton $\{\delta'(s, a)\}$, so a dummy state d_t is a supported successor in $\mathfrak{R}(\mathcal{G})$ exactly when it is a supported successor in $\mathcal{G}'$.

Now suppose the corresponding prefixes end at d_t. In $\mathcal{G}'$, the effective action $\top$ leads deterministically to $t \in S_{\mathsf{min}}$, after which the min-player uses $\lambda = \pi'(oh'_t)(t) \in \Delta(A_{\mathsf{min}})$. In $\mathfrak{R}(\mathcal{G})$, the mapped environment selects $\mathsf{bar}^{\mathcal{G}'}_t(\lambda)$ at $(d_t, \top)$. Hence, for every successor $s' \in S_{\mathsf{max}}$,

$$\mathsf{bar}^{\mathcal{G}'}_t(\lambda)(s') > 0 \quad \Longleftrightarrow \quad \lambda(a)\delta'(t, a)(s') > 0 \text{ for some } a \in A_{\mathsf{min}},$$

because all summands are nonnegative. Therefore an extension from d_t to s' is supported in $\mathfrak{R}(\mathcal{G})$ exactly when an extension from d_t through t and some supported min-action a to s' is supported in $\mathcal{G}'$. This proves the induction step for the first equality.

For the reverse direction, agent-action supports agree by the definition of Ξ_{max}. At a dummy state use $\lambda = \Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})(oh')(t)$. The first identity in Lemma C.6 gives $\mathsf{bar}^{\mathcal{G}'}_t(\lambda) = \pi^{\mathfrak{R}}(\Lambda^{\mathsf{min}}_{\mathcal{O}}(oh'), \top)(d_t, \top)$, so the same nonnegativity argument proves the reverse induction step. Finally, an infinite play is supported exactly when all its finite prefixes are supported. Applying Λ to the resulting prefix correspondences yields both displayed equalities. □

Lemma C.12. *Let $W' \subseteq \mathsf{Plays}_{\mathcal{G}'}$ be a parity objective of $\mathcal{G}'$ and $W^{\mathfrak{R}} = \Lambda(W')$ its lifted objective in $\mathfrak{R}(\mathcal{G})$. Let $\sigma \in \Sigma_{\mathcal{G}'}$ and $\sigma^{\mathfrak{R}} = \Omega_{\mathsf{max}}(\sigma)$. Then*

$$\sigma \text{ is sure winning for } W' \iff \sigma^{\mathfrak{R}} \text{ is sure winning for } W^{\mathfrak{R}}.$$

Proof. Suppose σ is sure winning and take arbitrary $\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}}(\mathcal{G})$. The second equality of Lemma C.11 gives

$$\mathsf{Plays}_{\mathfrak{R}}^{\sigma^{\mathfrak{R}},\pi^{\mathfrak{R}}}(\mathcal{G}) = \Lambda\Big(\mathsf{Plays}_{\mathcal{G}'}^{\sigma,\Xi_{\mathsf{min}}(\pi^{\mathfrak{R}})}\Big) \subseteq W^{\mathfrak{R}}.$$

Conversely, suppose $\sigma^{\mathfrak{R}}$ is sure winning and take arbitrary $\pi' \in \Pi_{\mathcal{G}'}$. The first equality gives $\Lambda(\mathsf{Plays}_{\mathcal{G}'}^{\sigma,\pi'}) = \mathsf{Plays}_{\mathfrak{R}}^{\sigma^{\mathfrak{R}},\Omega_{\mathsf{min}}(\pi')}(\mathcal{G}) \subseteq W^{\mathfrak{R}}$. Lemma C.7 then implies $\mathsf{Plays}_{\mathcal{G}'}^{\sigma,\pi'} \subseteq W'$. □

Lemma C.13 (Almost-sure Winning)**.** *Let $W' \subseteq \mathsf{Plays}_{\mathcal{G}'}$ be a parity objective of $\mathcal{G}'$ and $W^{\mathfrak{R}} = \Lambda(W')$ its lifted objective in $\mathfrak{R}(\mathcal{G})$. Let $\sigma \in \Sigma_{\mathcal{G}'}$ and $\sigma^{\mathfrak{R}} = \Omega_{\mathsf{max}}(\sigma)$. Then,*

$$\sigma \text{ is almost-sure winning for } W' \iff \sigma^{\mathfrak{R}} \text{ is almost-sure winning for } W^{\mathfrak{R}}.$$

Proof. The fixed-strategy equality proved in Lemma C.9 is

$$\inf_{\pi' \in \Pi_{\mathcal{G}'}} \mathbb{P}_{\mathcal{G}'}^{\sigma,\pi'}(W') = \inf_{\pi^{\mathfrak{R}} \in \Pi_{\mathfrak{R}}(\mathcal{G})} \mathbb{P}_{\mathfrak{R}}(\mathcal{G})^{\sigma^{\mathfrak{R}},\pi^{\mathfrak{R}}}(W^{\mathfrak{R}}).$$

Hence one side equals 1 if and only if the other does. □

Lemma C.14 (Limit-Sure Winning)**.** *Let $W' \subseteq \mathsf{Plays}_{\mathcal{G}'}$ be a parity objective of $\mathcal{G}'$ and $W^{\mathfrak{R}} = \Lambda(W')$ is the corresponding objective in $\mathfrak{R}(\mathcal{G})$. Then,*

$$W' \text{ is limit-sure} \iff W^{\mathfrak{R}} \text{ is limit-sure}$$

Proof. Follows directly from lemma C.9. □

Lemma C.15 (Quantitative Analysis)**.** *Let $W' \subseteq \mathsf{Plays}_{\mathcal{G}'}$ be a parity objective of $\mathcal{G}'$ and $W^{\mathfrak{R}} = \Lambda(W')$ is the corresponding objective in $\mathfrak{R}(\mathcal{G})$. Then, for $k \in (0,1)$*

$$\mathrm{Val}_{\mathcal{G}'}(W') \geq k \iff \mathrm{Val}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}}) \geq k$$

Proof. Follows directly from lemma C.9. □

C.3.1 Size and memory preservation between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$

Lemma C.16 (Size and memory preservation under RPOMDP reduction)**.** *Let $\mathcal{G}'$ be a Pre-min-transformed POSG and $\mathfrak{R}(\mathcal{G})$ its RPOMDP reduction. The maps $\Lambda : \mathsf{Plays}_{\mathcal{G}'} \to \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}$, $\Lambda_{\mathcal{O}}^{\mathsf{max}}, \Lambda_{\mathcal{O}}^{\mathsf{min}}$ on observation histories, and $\Omega_\star, \Xi_\star$ ($\star \in \{\mathsf{max}, \mathsf{min}\}$) on strategies satisfy the following.*

1. (size) $|S_{\mathfrak{R}}| = |S_{\mathsf{max}}| + |S_{\mathsf{min}}|$, $|A_{\mathfrak{R}}| = |A_{\mathsf{max}}| + 1$, *the total vertex count of $\mathcal{P}_{\mathfrak{R}}$ is at most $|S_{\mathsf{min}}|\,|A_{\mathsf{min}}| + |S_{\mathsf{max}}|\,|A_{\mathsf{max}}|$, and observation alphabets are inherited unchanged from $\mathcal{G}'$. In particular, $|\mathfrak{R}(\mathcal{G})\,| = \mathcal{O}(|\mathcal{G}'|)$ and the construction runs in linear time.*
2. (history length) *For every $oh' \in \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'}$, $|\Lambda_{\mathcal{O}}^{\mathsf{max}}(oh')| = \left\lceil \frac{2|oh'|+1}{3} \right\rceil$, and for every $oh^{\mathfrak{R}} \in \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})}$, $|(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh^{\mathfrak{R}})| = |oh^{\mathfrak{R}}| + \lfloor |oh^{\mathfrak{R}}|/2 \rfloor$. The same bounds hold for $\Lambda_{\mathcal{O}}^{\mathsf{min}}$ on $\mathsf{OH}_{\mathsf{min}}^{\mathcal{G}'}$ and $\widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}$. Observation histories on the two sides are linear in each other.*
3. (memory) *The lifting maps $\Omega_\star, \Xi_\star$ preserve memory class exactly: memoryless, finite-memory and infinite-memory regimes correspond exactly between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$.*

Proof. **Clause (1).** Direct from the construction. The state space deletes the min-states S_{min} retaining only $S_{\mathsf{max}} \cup D$, so $|S_{\mathfrak{R}}| = |S_{\mathsf{max}}| + |S_{\mathsf{min}}|$. The action set deletes the min-actions retaining only $A_{\mathsf{max}} \cup \{\top\}$, so $|A_{\mathfrak{R}}| = |A_{\mathsf{max}}| + 1$. The observation alphabets are $\mathcal{O}_{\mathsf{max}}^{\mathfrak{R}} = \mathcal{O}_{\mathsf{max}} \cup \{o_x^\dagger \mid x \in \mathcal{O}_{\mathsf{max}}\} = \mathcal{O}'_{\mathsf{max}}$ and $\mathcal{O}_{\mathsf{min}}^{\mathfrak{R}} = \mathcal{O}_{\mathsf{min}} \cup \{o_x^\# \mid x \in \mathcal{O}_{\mathsf{min}}\} = \mathcal{O}'_{\mathsf{min}}$, which have the same cardinalities as in $\mathcal{G}'$. For every $(s,a) \in S_{\mathsf{max}} \times A_{\mathsf{max}}$, the non-default vertex set $V_{s,a}$ is the singleton $\{\delta'(s,a)\}$.

For each $(d_s, \top)$, the vertex list consists of $\delta'(s,a)$ for all $a \in A_{\mathsf{min}}$. Hence the construction lists at most $|S_{\mathsf{max}}|\,|A_{\mathsf{max}}| + |S_{\mathsf{min}}|\,|A_{\mathsf{min}}|$ generators; all other globally enabled pairs share one default max-losing rule.

Each component is therefore a constant or unit-additive perturbation of the corresponding component in $\mathcal{G}'$, so $|\mathfrak{R}(\mathcal{G})\,| = \mathcal{O}(|\mathcal{G}'|)$ and $\mathfrak{R}(\mathcal{G})$ is computed from $\mathcal{G}'$ in linear time.

Clause (2). Let $oh' = o_0, z_1, o_1, o_2, z_3, o_3, \cdots, o_{2n-2}, z_{2n-1}, o_{2n-1}, o_{2n} \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$ ending in max state, so that $|oh'| = 3n+1$. By construction,

$$\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh') = o_0, z_1, \cdots, o_{2n},$$

where o_{2k+1} are removed for every 3 states. Hence

$$|\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh')| = 2n+1 = \tfrac{2|oh'|+1}{3}.$$

Let $oh' = o_0, z_1, o_1, o_2, z_3, o_3, \cdots, o_{2n-2}, z_{2n-1}, o_{2n-1}, o_{2n}, z_{2n+1} \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$ ending in dummy state, so that $|oh'| = 3n+2$. By construction,

$$\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh') = o_0, z_1, \cdots, o_{2n}, z_{2n+1}$$

where o_{2k+1} are removed for every 3 states. Hence

$$|\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh')| = 2n+2 = \tfrac{2|oh'|+2}{3}.$$

From above values, For every $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$, $|\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh')| = \left\lceil \frac{2|oh'|+1}{3} \right\rceil$

Let $oh = o_0, z_1, \cdots, o_{2n} \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{a}$ ending in S_{max} state, so that $|oh| = 2n+1$. By construction,

$$(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh) = o_0, z_1, o_1, o_2, z_3, o_3, \cdots, o_{2n-2}, z_{2n-1}, o_{2n-1}, o_{2n}$$

where o_{2k+1} are added deterministically from z_{2k+1} for every 2 states. Hence

$$|(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh)| = 3n+1 = \tfrac{3|oh|-1}{2} = |oh| + \tfrac{|oh|-1}{2}.$$

Let $oh = o_0, z_1, \cdots, o_{2n}, z_{2n+1} \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{a}$ ending in dummy state, so that $|oh| = 2n+2$. By construction,

$$(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh) = o_0, z_1, o_1, o_2, z_3, o_3, \cdots, o_{2n-2}, z_{2n-1}, o_{2n-1}, o_{2n}, z_{2n+1}$$

where o_{2k+1} are added from z_{2k+1} for every 2 states. Hence

$$|(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh)| = 3n+2 = \tfrac{3|oh|}{2} - 1 = |oh| + \tfrac{|oh|-2}{2}.$$

Thus, for every $oh \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}}$, $|(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh)| = |oh| + \left\lfloor \frac{|oh|-1}{2} \right\rfloor$. The environment-history bounds follow identically.

Clause (3). Fix a player and let its strategy be realized by a transducer $M = (Q, q_0, u, g)$ over that player's observation alphabet. The history maps $\Lambda^{\mathsf{max}}_{\mathcal{O}}$ and $\Lambda^{\mathsf{min}}_{\mathcal{O}}$ delete each S_{min}-observation, while their inverses reinsert it. The deleted observation is determined by the preceding tagged dummy observation: $o^{\dagger}_x$ or $o^{\#}_x$ uniquely determines x. Consequently, both contraction and expansion can be performed online without adding persistent memory states.

For Ω_{max}, run the transducer for σ' on the expanded history $(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}})$ and use its state-indexed action assignment unchanged. For Ω_{min}, run the transducer for π' on $(\Lambda^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}})$ and apply $\mathsf{bar}^{\mathcal{G}'}_{s}$ to the output distribution at a compatible dummy state; at a non-dummy state the required transition assignment is the unique element of the singleton uncertainty set. These are output transformations and require no additional memory.

Conversely, Ξ_{max} runs the transducer for $\sigma^{\mathfrak{R}}$ on the contracted history $\Lambda^{\mathsf{max}}_{\mathcal{O}}(oh')$. The map Ξ_{min} runs the transducer for $\pi^{\mathfrak{R}}$ on $\Lambda^{\mathsf{min}}_{\mathcal{O}}(oh')$ and applies the fixed coefficient choice β^{P}_{s} to the selected polytope point. Again, contraction and the output transformation leave the memory set Q unchanged. The history identities in Lemmas C.3 and C.4 show that these transducers realize the four stated strategy maps.

Thus each map preserves the number of memory states. In particular, memoryless, finite-memory, and unrestricted infinite-memory strategies correspond in both directions. □

C.4 Proof of Lemma 3.3

Proof. Because parity objective subsumes all ω-regular objectives, from lemma C.9, $\mathrm{Val}_{\mathcal{G}'}(W') = \mathrm{Val}_{\mathfrak{R}(\mathcal{G})}(W^{\mathfrak{R}})$ for any ω-regular objective W' in $\mathcal{G}'$ and its corresponding objective $W^{\mathfrak{R}}$ in $\mathfrak{R}(\mathcal{G})$. The "consequently" statement follows from lemmas C.12, C.13, C.14, and C.15. □