SOLVING ROBUST POMDPS WITH ω-REGULAR OBJECTIVES VIA PARTIALLY OBSERVABLE STOCHASTIC GAMES

Anonymous authors
Paper under double-blind review

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 (Kaelbling et al., 1998). 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 (Chatterjee et al., 2024). In the fully observable setting, this challenge is addressed through robust MDPs (RMDPs) (Nilim & Ghaoui, 2003; Iyengar, 2005), 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) (Osogami, 2015) provide a robust generalization of POMDPs.

Recent years have seen significant advances in algorithmic approaches to solving RMDPs and RPOMDPs. Existing works largely focus on reward objectives, such as maximizing the expected discounted-sum (Nilim & Ghaoui, 2003; Iyengar, 2005; Wiesemann et al., 2013; Kaufman & Schaefer, 2013; Tamar et al., 2014; Ghavamzadeh et al., 2016; Ho et al., 2021) or the long-run average reward (Tewari & Bartlett, 2007; Grand-Clément & Petrik, 2023; Goyal & Grand-Clément, 2023; Wang et al., 2023; Chatterjee et al., 2024; Meggendorfer et al., 2025; Asadi et al., 2026a) in the R(PO)MDP, which optimize for performance-related properties in the model. 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 (Summers & Lygeros, 2010). The stability correctness property requires the agent to eventually reach and stay within some target region ad infinitum (Khalil & Grizzle, 2002). 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) (Pnueli, 1977), or more generally ω-regular specifications (Clarke et al., 2018), which subsumes most logical correctness properties appearing in control and robotics

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

applications (Kress-Gazit et al., 2009). 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 (Chatterjee et al., 2008; Sen et al., 2006), or is restricted to probability 1 or limit-sure satisfaction of ω-regular specifications (Asadi et al., 2026b).

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 (Givan et al., 2000) or L^1-uncertainty sets (Ho et al., 2021). Our approach is based on the following fundamental novel result: we formally prove that the problem of solving RPOMDPs (defined in section 2) with ω-regular objectives is *equivalent* to the problem of solving partially observable stochastic games (POSGs) (defined in section 3) 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 (Chatterjee et al., 2024; Grand-Clément et al., 2023; Bovy et al., 2024). Here, we consider ω-regular objectives rather than reward objectives. Moreover, we also present a reduction in the opposite direction from POSGs to RPOMDPs, which is technically more challenging. Since solving POSGs under ω-regular objectives has received more attention (Chatterjee et al., 2013) compared to RPOMDPs, our equivalence result allows us to derive a plethora of new complexity results on solving R(PO)MDPs. 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. Our equivalence result holds for Sure Winning, Almost Sure Winning, Limit Sure Winning, and Quantitative Winning ω-regular objectives.
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 and one-sided partial observability setting (1s RPOMDP).

Thus, we significantly advance 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.

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 (Roy et al., 2017; Tessler et al., 2019; Wang & Zou, 2021; 2022; Yu et al., 2023; Wang et al., 2024). Algorithmic approaches for solving RPOMDPs under reward objectives were considered in (Osogami, 2015; Chamie & Mostafa, 2018; Suilen et al., 2020; Nakao et al., 2021; Cubuktepe et al., 2021; Galesloot et al., 2025a;b; Krale et al., 2025). See (Suilen et al., 2024) 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 the structure of uncertainty sets such as interval MDPs (Chatterjee et al., 2008; Sen et al., 2006), or is restricted to probability 1 or limit-sure satisfaction of ω-regular specifications (Asadi et al., 2026b) where the two objectives are shown to have the same value, but without exact complexity bounds that we establish.

The existence of a reduction from R(PO)MDPs to (PO)SGs under reward objectives has been observed in (Nilim & Ghaoui, 2003; Tessler et al., 2019; Goyal & Grand-Clément, 2023; Grand-Clément et al., 2023). The work (Chatterjee et al., 2024) formalizes the reduction from RMDPs to turn-based stochastic games under both expected discounted-sum and long-run average objectives. The work (Bovy et al., 2024) formalizes the reduction from RPOMDPs to POSGs under expected cumulative reward and discounted-sum objectives. However, none of these consider ω-regular objectives.

	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 (Chatterjee et al., 2013). The table considers RMDPs (full observability for both the agent and the environment), 1s RPOMDPs (partial observability for the agent only), and two-sided 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 (Chatterjee et al., 2013). The table numbers correspond to the tables in (Chatterjee et al., 2013) from which the respective complexity results on POSGs were obtained. Turn-based quantitative stochastic parity games complexity results, from which we infer complexity results for quantitative analysis of RMDPs, were also derived in (Chatterjee et al., 2004).

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 $\mathrm{last}(b) = b_n$.

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 RPOMDPs (Nilim & Ghaoui, 2003; Iyengar, 2005), 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 transition probabilities independently across state-action pairs. Moreover, we assume that each $P(s,a)$ is a *polytope* with a finite vertex set $V_{s,a}$. $P(s,a) = \{\sum_{i=1}^{|V_{s,a}|} \alpha_i v_i \mid \alpha_i \in [0,1], \sum_i \alpha_i = 1, v_i \in V_{s,a}\}$, where each vertex $v \in V_{s,a}$ has the form $(x_1, \ldots, x_{|S|})$ with $x_i \geq 0$ and $\sum_i x_i = 1$. 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-uncertainty sets and interval (PO)MDPs (Ho et al., 2021; Givan et al., 2000). 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 (the environment) observation history and a_t, the adversarial environment selects a complete transition assignment, namely, a $|S| \times |A|$ vector $P_t \in \mathcal{P}$; the transition from the actual state uses its component $P_t(s_t, a_t) \in \mathcal{P}(s_t, a_t)$ ($\mathcal{P}(s, a)$ denotes the (s, a)th component of $\mathcal{P}$). The environment observes only $o_t^e = \mathcal{Z}_{\mathsf{e}}(s_t)$. 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 (Bovy et al., 2024), our semantics model *dynamic uncertainty* where the environment may select a new transition assignment at each time step.

A *history* is a finite sequence $h_t = s_0, (o_0^a, o_0^e), s_1, \ldots, s_t, (o_t^a, o_t^e) \in (S \times \mathcal{O}_{\mathsf{a}} \times \mathcal{O}_{\mathsf{e}})^*$. An *agent* (resp., *environment*) *action invisible observation history* for h_t is a sequence $oh_t^a = o_0^a, \ldots, o_t^a \in (\mathcal{O}_{\mathsf{a}})^* \times \mathcal{O}_{\mathsf{a}}$ (resp., $oh_t^e = o_0^e, \ldots, o_t^e \in (\mathcal{O}_{\mathsf{e}})^* \times \mathcal{O}_{\mathsf{e}}$). Let OH_a and OH_e denote the sets of agent and environment observation histories. An *agent strategy* and an *environment strategy* are mappings $\sigma : \mathsf{OH}_a \to \Delta(A), \pi : \mathsf{OH}_e \times A \to \mathcal{P}$, respectively. Thus $\sigma(oh)(a)$ is the probability of choosing action a after observation history oh, independent of the hidden current state. The environment still chooses a complete transition assignment in response to the action played: $\pi(oh, a)$ is a tuple with one component $\pi(oh, a)(s, a) \in P_{s,a}$ for every state s compatible with oh, since the local uncertainty sets differ from state to state.

Given a RPOMDP $\mathcal{M}$, let $\Sigma_{\mathcal{M}}$ and $\Pi_{\mathcal{M}}$ denote all agent and environment strategies. 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 action distribution or transition 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, the agent selects $\alpha_t = \sigma(oh_t^a) \in \Delta(A)$ and samples a_t from α_t. The environment then selects $P_t = \pi(oh_t^e, a_t) \in \mathcal{P}$. The next state satisfies $\Pr(s_{t+1} \mid h_t, a_t) = P_t(s_t, a_t)(s_{t+1})$. Evaluating the current-state component of P_t does not reveal that state to the environment. 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 $\mathbb{P}_{s_0}^{\sigma,\pi}(\cdot)$ over the set $\mathsf{Plays}_{\mathcal{M}}$ (Billingsley, 2017). $\mathsf{Plays}_{\mathcal{M}}^{\sigma,\pi}$ denotes 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 (Clarke et al., 2018), 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 (Thomas, 1997). Parity objectives in RPOMDP are realised by equipping them 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.

Throughout, we make the standard assumption in RPOMDPs (and later in ac-POSGs introduced in Section 3) 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. In figures, we only display effective actions and edges; every omitted state-action pair is implicitly completed by an absorbing outcome losing for its owner. This completion is without loss of generality.

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 $\mathrm{Val}_{\mathcal{M}}(W) \geq p$.

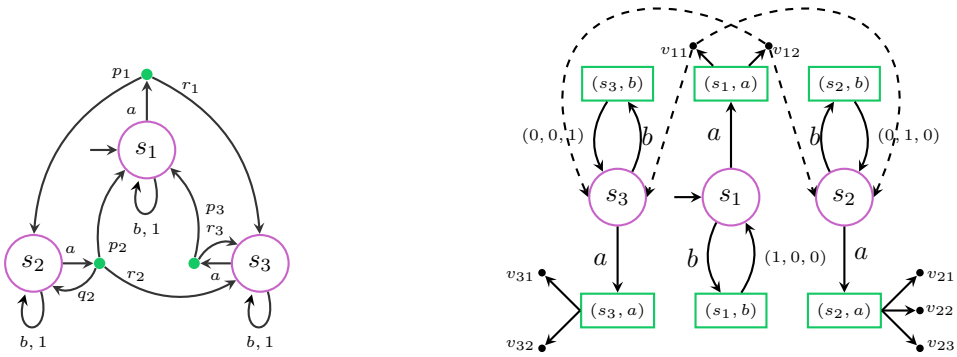

Figure 1: (a) RPOMDP $\mathcal{M}$ with states s_1, s_2, s_3 and actions a, b. Each transition arrow consists of two parts, the first denotes an action (a or b) whereas the second denotes the probability of moving to a state upon taking that action (e.g. p_1 is the probability of moving to s_2 upon taking 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.

3 Equivalence with Partially Observable Stochastic Games

This section presents the main result of the paper. We provide reductions between RPOMDPs and alternating control POSGs for each problem above. Both reductions are linear in the model size.

Partially Observable Stochastic Games (POSGs). A turn-based, two-player (max, min) *POSG* (Chatterjee et al., 2013) is a tuple $\mathcal{G} = (S_{\max}, S_{\min}, A_{\max}, A_{\min}, \delta, s_0, \mathcal{O}_{\max}, \mathcal{O}_{\min}, \mathcal{Z}_{\max}, \mathcal{Z}_{\min})$, where $S_{\max}$ and $S_{\min}$ are the sets of max- and min-states, respectively. Write $S^{\mathcal{G}} = S_{\max} \uplus S_{\min}$, 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 $A^{\mathcal{G}} = A_{\max} \cup A_{\min}$ and $A_{\max}(s)$ (resp. $A_{\min}(s)$) denote the set of effective actions at $s \in S_{\max}$ (resp. $s \in S_{\min}$). The transition function is $\delta : (S_{\max} \times A_{\max}) \cup (S_{\min} \times A_{\min}) \to \Delta(S^{\mathcal{G}})$. Observations are given by sets $\mathcal{O}_{\max}, \mathcal{O}_{\min}$ and functions $\mathcal{Z}_{\max} : S^{\mathcal{G}} \to \mathcal{O}_{\max}$, $\mathcal{Z}_{\min} : S^{\mathcal{G}} \to \mathcal{O}_{\min}$. A POSG is *alternating control (ac-POSG)* if $\delta(s, a) \in \Delta(S_{\min})$ for every $(s, a) \in S_{\max} \times A_{\max}$ and $\delta(s, a) \in \Delta(S_{\max})$ for every $(s, a) \in S_{\min} \times A_{\min}$. Throughout, ac-POSGs start in a max-state, $s_0 \in S_{\max}$; otherwise, a fresh deterministic initial max-state can be prepended.

A *history* of $\mathcal{G}$ is a sequence $h_t = s_0, (o_0^1, o_0^2), s_1, \ldots, s_t, (o_t^1, o_t^2)$ in $(S^{\mathcal{G}} \times \mathcal{O}_{\max} \times \mathcal{O}_{\min})^*$ such that for all $0 \le i \le t$, $o_i^1 = \mathcal{Z}_{\max}(s_i), o_i^2 = \mathcal{Z}_{\min}(s_i)$. Moreover, for every $0 \le i < t$, there is an action $a \in A_{\max}$ if $s_i \in S_{\max}$, or $a \in A_{\min}$ if $s_i \in S_{\min}$, such that $\delta(s_i, a)(s_{i+1}) > 0$. For $\star \in \{\max, \min\}$, if h_t ends in an $S_\star$-state, its *$\star$-(action invisible) 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. For $o \in \mathcal{O}_{\min}$, define the *observation class* of o as $S_{\min}(o) = \{s \in S_{\min} : \mathcal{Z}_{\min}(s) = o\}$. A max-player strategy σ maps each observation history to a single action distribution. The min-player uses *factored prescriptions* π: for an observation o, a prescription chooses one local action distribution for every min-state in $S_{\min}(o)$. Thus, we define max-player strategy $\sigma : \mathsf{OH}_{\max}^{\mathcal{G}} \to \Delta(A_{\max})$ and min-player strategy $\pi : \mathsf{OH}_{\min}^{\mathcal{G}} \to \bigcup_{o \in \mathcal{O}_{\min}} \prod_{s \in S_{\min}(o)} \Delta(A_{\min})$, where $\pi(oh)$ belongs to the product for $o = \mathrm{last}(oh)$ and each component is supported on the effective min-actions at its state (Section 2).

Plays induced by (σ, π) are defined analogously to RPOMDP plays, and $\mathsf{Plays}_{\mathcal{G}}$ denotes the set of all such plays. Probability measures over plays, priority functions, objectives $W \subseteq \mathsf{Plays}_{\mathcal{G}}$, the value $\mathrm{Val}_{\mathcal{G}}(W)$, and the four analysis problems are all defined exactly as for RPOMDPs. Let $\Sigma_{\mathcal{G}}$

and $\Pi_{\mathcal{G}}$ denote the sets of all max- and min-strategies, respectively. The computational analysis of ordinary observation-based POSGs was studied by Chatterjee et al. (2013), whose min-strategies $\widehat{\pi} : \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}} \to \Delta(A_{\mathsf{min}})$ assign a single distribution to each observation history. Our factored prescriptions generalize this notion: Lemmas A.3 and A.4 compare these notions, and the fact that this *does not affect our complexity results* is established at the end of Section 3.2.2.

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 (Chatterjee et al., 2024). 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 $\mathcal{G} = (S^{\mathcal{G}}_{\mathsf{max}}, S^{\mathcal{G}}_{\mathsf{min}}, A^{\mathcal{G}}_{\mathsf{max}}, A^{\mathcal{G}}_{\mathsf{min}}, \delta^{\mathcal{G}}, s^{\mathcal{G}}_0, \mathcal{O}^{\mathcal{G}}_{\mathsf{max}}, \mathcal{O}^{\mathcal{G}}_{\mathsf{min}}, \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}, \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}})$ and the corresponding objective. The formal construction is deferred to Appendix B.

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. 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.

States: max-player states $S^{\mathcal{G}}_{\mathsf{max}} = S$ are the states of the original RPOMDP, min-player states $S^{\mathcal{G}}_{\mathsf{min}}$ are all state–action pairs $(s, a) \in S \times A$ with a effective at s, and the initial state is $s^{\mathcal{G}}_0 = s_0 \in S^{\mathcal{G}}_{\mathsf{max}}$. *Actions:* The max-player's actions are those of the RPOMDP, $A^{\mathcal{G}}_{\mathsf{max}} = A$, with the same effective support at s. At min-state (s, a) the effective actions are $V_{s,a}$, and the global min-action alphabet is $A^{\mathcal{G}}_{\mathsf{min}} = \bigcup_{(s,a)\in S^{\mathcal{G}}_{\mathsf{min}}} V_{s,a}$. Every min-state is supported on its local vertex set. *Observations:* $\mathcal{O}^{\mathcal{G}}_{\mathsf{max}} = \mathcal{O}_{\mathsf{a}}$ and $\mathcal{O}^{\mathcal{G}}_{\mathsf{min}} = \mathcal{O}_{\mathsf{e}}$, with $\mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}(s) = \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}((s, a)) = \mathcal{Z}_{\mathsf{a}}(s)$ and $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s) = \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}((s, a)) = \mathcal{Z}_{\mathsf{e}}(s)$. *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 $v \in V_{s,a}$, $\delta^{\mathcal{G}}(s', v) = v$ is the distribution over $S^{\mathcal{G}}_{\mathsf{max}}$ specified by the vertex v of the uncertainty polytope $P(s, a)$. Randomizing over min-actions therefore produces every distribution in $P(s, a)$.

POSG 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 B.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 is deterministic 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 copies of all min-player's states, and $\mathcal{O}^\dagger = \{o_x^\dagger \mid x \in \mathcal{O}_{\mathsf{max}}\}$ 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 $A'_{\mathsf{min}} = A_{\mathsf{min}}$. Original states retain their effective alphabets $A_{\mathsf{max}}(s)$ and $A_{\mathsf{min}}(s)$, while dummy states have sole effective action $\top$.
- *Transition function.* (i) For $s_1 \in S_{\mathsf{max}}$, $a_1 \in A_{\mathsf{max}}(s_1)$ and $s_2 \in S_{\mathsf{min}}$, $\delta'(s_1, a_1)(d_{s_2}) = \delta(s_1, a_1)(s_2)$. (ii) For $s_2 \in S_{\mathsf{min}}$ and $a_2 \in A_{\mathsf{min}}(s_2)$, set $\delta'(s_2, a_2) = \delta(s_2, a_2)$. (iii) At a dummy state set $\delta'(d_s, \top) = \mathrm{Dirac}(s)$.
- *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.

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 s_{2k} and the S_{min}-step s_{2k+1}. 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', strips the inserted observations and queries σ at the underlying $\mathcal{G}$-history: $\Phi_\star(\sigma)(oh') = \sigma((\Gamma^\star_O)^{-1}(oh'))$; at dummy states $\Phi_{\mathsf{max}}(\sigma)$ plays $\top$ deterministically. For max this equality copies one action distribution; for min it copies the complete prescription. For $\mathcal{G}'$-strategy σ', 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, $|\Gamma^\star_O(oh)|$ and $|(\Gamma^\star_O)^{-1}(oh')|$ are linear in $|oh|$ and $|oh'|$, respectively. 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 C.3.1.

Lemma 3.2 (Equivalence between $\mathcal{G}$ and $\mathcal{G}'$, Appendix C.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 $\mathcal{G}$, objective W and its corresponding 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 polynomial 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 $s_1 \xrightarrow{a_1} d_{s_2} \xrightarrow{\top} s_2 \xrightarrow{a_2} s_3 \xrightarrow{a_3} d_{s_4} \xrightarrow{\top} s_4 \cdots$, in which each round has the order $S_{\mathsf{max}}, D, S_{\mathsf{min}}$. The key observation is that the min-player's only role is to pick an action $a \in A_{\mathsf{min}}(s)$ 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: $P(d_s, \top)$ is the convex hull of the distributions $\delta'(s, a)$ for $a \in A_{\mathsf{min}}(s)$. Picking a point in this

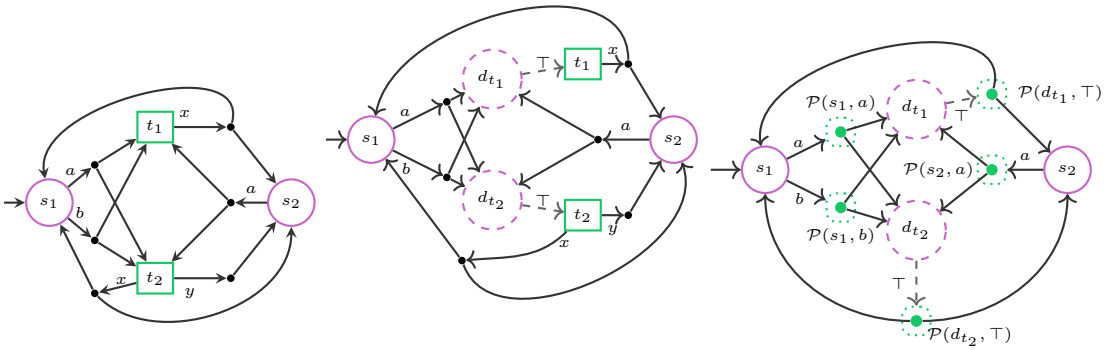

Figure 2: (a) A POSG $\mathcal{G}$. max-states s_1, s_2 are circles, while min-states t_1, t_2 are rectangles. (b) The Pre-min transformed game $\mathcal{G}'$ with dashed circles representing new dummy states- d_{t_1}, d_{t_2} and dashed arrows represent the $\top$-action. (c) The reduced RPOMDP $\mathfrak{R}(\mathcal{G})$ with red dotted circles representing 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 $\{\delta'(t_1, x)\}$. $P(d_{t_2}, \top)$ is the polytope with vertex set $\{\delta'(t_2, x), \delta'(t_2, y)\}$.

polytope is a convex combination $\sum_a \alpha_a \, \delta'(s, a)$, where α_a is the distribution value induced by $\mathsf{min}-$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 effective alphabets $A_{\mathfrak{R}}(s) = A_{\mathsf{max}}(s)$ at original max-states and $A_{\mathfrak{R}}(d_s) = \{\top\}$.
- *Uncertainty set.* $\mathcal{P}_{\mathfrak{R}} = \prod_{(s,a) \in S_{\mathfrak{R}} \times A_{\mathfrak{R}}} P(s, a)$. The effective polytopes are generated by $V_{s,a} = \{\delta'(s, a)\}$ for $s \in S_{\mathsf{max}}$, $a \in A_{\mathsf{max}}(s)$, and $V_{d_{s'}, \top} = \{\delta'(s', a') : a' \in A_{\mathsf{min}}(s')\}$.
- *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'. Unlike Γ in Section 3.2.1, 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 $\Lambda^{\mathsf{max}}_{\mathcal{O}}, \Lambda^{\mathsf{min}}_{\mathcal{O}}$: 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 environment map $\Lambda^{\mathsf{min}}_{\mathcal{O}}$ is a bijection onto 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 $\sum_{a \in A_{\mathsf{min}}(s')} \alpha_a \cdot \delta'(s', a) \in P(d_{s'}, \top)$. 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 uses a fixed choice of convex coefficients over the effective min-actions. For $\star \in \{\mathsf{max}, \mathsf{min}\}$, $\Omega_\star, \Xi_\star$ are inverses of each other by the observation-history correspondence. The formal maps and proofs appear in Appendix D.1.

• **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 D.3.1).

Lemma 3.3 (Equivalence between $\mathcal{G}'$ and $\mathfrak{R}(\mathcal{G})$, Proof in Appendix D.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) $\Lambda(\mathsf{Plays}_{\mathcal{G}'}^{\sigma',\pi'}) = \mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}$;
(III, prob.) $\mathbb{P}_{\mathcal{G}'}^{\sigma',\pi'}(W') = \mathbb{P}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}(W^{\mathfrak{R}})$; (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. (I) Λ leaves the underlying $S_{\mathsf{max}} \cup D$-state sequence, and hence the positions of target-state visits unchanged, and $p^{\mathfrak{R}}$ is just p' restricted to $S_{\mathfrak{R}}$; neither change affects the $\liminf$ the parity condition depends on. (II) Following $\rho' \in \mathsf{Plays}_{\mathcal{G}'}^{\sigma',\pi'}$ step by step, $\Lambda(\rho')$ is producible by $(\Omega_{\mathsf{max}}(\sigma'), \Omega_{\mathsf{min}}(\pi'))$ in $\mathfrak{R}(\mathcal{G})$: at each S_{max}-state the singleton uncertainty set forces the same transition, and at each dummy state the convex combination $\sum_a \alpha_a \delta'(s, a)$ chosen by $\Omega_{\mathsf{min}}(\pi')$ is exactly the distribution π' induces at s. Conversely, every play in $\mathsf{Plays}_{\mathfrak{R}(\mathcal{G})}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}$ has a Λ-preimage in $\mathcal{G}'$, obtained by drawing a min-action at each dummy step according to the coefficients α_a. (III) A cylinder-set induction on $\mathcal{G}'$-history, splitting each step into a max-action reaching a dummy state and the forced dummy step followed by the min-action reaching the next max-state; both cases reduce the $\mathfrak{R}(\mathcal{G})$-extension probability to the matching $\mathcal{G}'$-probability. The converse follows the same way by induction on $\mathfrak{R}(\mathcal{G})$-history. (IV) applies (III) in both directions to equate the inner infima for each fixed agent strategy, then takes the outer supremum. *Consequently:* quantitative and limit-sure analysis follow from (IV); almost-sure follows from (III), taking the infimum over π' on each side and matching witnesses through Ω, Ξ; sure analysis follows from (II) and (I), since a strategy keeping all $\mathcal{G}'$-plays inside W' corresponds via Ω to one keeping all $\mathfrak{R}(\mathcal{G})$-plays inside $W^{\mathfrak{R}}$, and conversely via Ξ. From Lemmas 3.2 and 3.3, we conclude:

Theorem 3.4. *For an ac-POSG $\mathcal{G}$ and its RPOMDP construction $\mathfrak{R}(\mathcal{G})$ (Section 3.2), the sure, almost-sure, limit-sure, and quantitative analyses of any ω-regular objective are inter-reducible in polynomial time between $\mathcal{G}$ and $\mathfrak{R}(\mathcal{G})$.*

Complexity Results. The reductions in Sections 3.1 and 3.2 establish equivalence between (s, a) rectangular polytopic RPOMDPs and ac-POSGs as defined in sections 2 and 3. Lemma A.3 compiles the ac-POSGs in Section 3 into ordinary ac-POSG Chatterjee et al. (2013) with values and all four objectives preserved with polynomial overhead; together with Section 3.1, upper bounds in Chatterjee et al. (2013) apply to our RPOMDPs. Lemma A.4 converts ordinary ac-POSGs into the class-constant subclass (the same distribution assigned to every min-state of an observation class) of our ac-POSGs at no cost with values and all four objectives preserved; together with Section 3.2, the lower bounds from Chatterjee et al. (2013) apply to our RPOMDPs. Allowing full observability to the min-player, factored prescriptions coincide with ordinary min strategies of Chatterjee et al. (2013). In this case, the reductions of 3.1 and 3.2, show the equivalence of 1s RPOMDPs and one sided ordinary ac-POSGs. Thanks to our unified treatment across observability regimes, when both players have full observation, we have the equivalence between stochastic games and RMDPs, allowing to lift complexity results of Chatterjee et al. (2013) to 1s RPOMDPs and RMDPs.

Concluding Remarks. We studied RPOMDPs under ω-regular objectives, and showed that solving them is equivalent to solving POSGs under ω-regular objectives. We do this by establishing polynomial-time reductions in both directions. Exploiting this equivalence, we derived several new complexity results for both RPOMDPs and RMDPs, and provided a complete picture of the complexity landscape of the problem. Interesting future work would be to consider lifting rectangularity and polytopic uncertainty set assumptions, and studying complexity under these more general settings.

AI USE STATEMENT

Generative AI tools assisted with writing concisely, drawing figures and checking for consistency in notations. The authors are responsible for independently verifying all mathematical claims, constructions, and proofs before submission, and for ensuring that this disclosure accurately reflects the assistance used. We take responsibility for the final content of this work.

REFERENCES

Ali Asadi, Krishnendu Chatterjee, Ehsan Kafshdar Goharshady, Mehrdad Karrabi, Alipasha Montaseri, and Carlo Pagano. Strongly polynomial time complexity of policy iteration for l_∞ robust mdps. *CoRR*, abs/2601.23229, 2026a. doi: 10.48550/ARXIV.2601.23229. URL https://doi.org/10.48550/arXiv.2601.23229.

Ali Asadi, Krishnendu Chatterjee, Ehsan Kafshdar Goharshady, Mehrdad Karrabi, and Ali 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*, pp. 36137–36145, 2026b. doi: 10.1609/AAAI.V40I43.40931. URL https://doi.org/10.1609/aaai.v40i43.40931.

Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2017.

Eline M. Bovy, Marnix Suilen, Sebastian Junges, and Nils 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*, pp. 6697–6706, 2024. URL https://www.ijcai.org/proceedings/2024/740.

Mahmoud El Chamie and Hala 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*, pp. 5586–5591, 2018. doi: 10.1109/CDC.2018.8619468. URL https://doi.org/10.1109/CDC.2018.8619468.

Krishnendu Chatterjee, Marcin Jurdzinski, and Thomas A. Henzinger. Quantitative stochastic parity games. In J. Ian Munro (ed.), *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*, pp. 121–130. SIAM, 2004. URL http://dl.acm.org/citation.cfm?id=982792.982808.

Krishnendu Chatterjee, Koushik Sen, and Thomas 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*, pp. 302–317, 2008. doi: 10.1007/978-3-540-78499-9_22. URL https://doi.org/10.1007/978-3-540-78499-9_22.

Krishnendu Chatterjee, Laurent Doyen, and Thomas A. Henzinger. A survey of partial-observation stochastic parity games. *Formal Methods Syst. Des.*, 43(2):268–284, 2013. doi: 10.1007/S10703-012-0164-2. URL https://doi.org/10.1007/s10703-012-0164-2.

Krishnendu Chatterjee, Ehsan Kafshdar Goharshady, Mehrdad Karrabi, Petr Novotný, and Dorde 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*, pp. 6707–6715, 2024. URL https://www.ijcai.org/proceedings/2024/741.

Edmund M. Clarke, Thomas A. Henzinger, Helmut Veith, and Roderick Bloem (eds.). *Handbook of Model Checking*. Springer, 2018. ISBN 978-3-319-10574-1. doi: 10.1007/978-3-319-10575-8. URL https://doi.org/10.1007/978-3-319-10575-8.

Murat Cubuktepe, Nils Jansen, Sebastian Junges, Ahmadreza Marandi, Marnix Suilen, and Ufuk 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, pp. 11792–11800, 2021. doi: 10.1609/AAAI.V35I13.17401. URL https://doi.org/10.1609/aaai.v35i13.17401.

Maris F. L. Galesloot, Roman Andriushchenko, Milan Ceska, Sebastian Junges, and Nils 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*, pp. 8518–8526, 2025a. doi: 10.24963/IJCAI.2025/947. URL https://doi.org/10.24963/ijcai.2025/947.

Maris F. L. Galesloot, Marnix Suilen, Thiago D. Simão, Steven Carr, Matthijs T. J. Spaan, Ufuk Topcu, and Nils 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)*, pp. 4823–4831, 2025b. doi: 10.3233/FAIA251391. URL https://doi.org/10.3233/FAIA251391.

Mohammad Ghavamzadeh, Marek Petrik, and Yinlam 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*, pp. 2298–2306, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/9a3d458322d70046f63dfd8b0153ece4-Abstract.html.

Robert Givan, Sonia M. Leach, and Thomas L. Dean. Bounded-parameter markov decision processes. *Artif. Intell.*, 122(1-2):71–109, 2000. doi: 10.1016/S0004-3702(00)00047-3. URL https://doi.org/10.1016/S0004-3702(00)00047-3.

Vineet Goyal and Julien Grand-Clément. Robust markov decision processes: Beyond rectangularity. *Math. Oper. Res.*, 48(1):203–226, 2023. doi: 10.1287/MOOR.2022.1259. URL https://doi.org/10.1287/moor.2022.1259.

Julien Grand-Clément and Marek 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. URL http://papers.nips.cc/paper_files/paper/2023/hash/a4e720ce31ccd8ba747d8863e1580fa8-Abstract-Conference.html.

Julien Grand-Clément, Marek Petrik, and Nicolas Vieille. Beyond discounted returns: Robust markov decision processes with average and blackwell optimality. *CoRR*, abs/2312.03618, 2023. doi: 10.48550/ARXIV.2312.03618. URL https://doi.org/10.48550/arXiv.2312.03618.

Chin Pang Ho, Marek Petrik, and Wolfram Wiesemann. Partial policy iteration for l1-robust markov decision processes. *J. Mach. Learn. Res.*, 22:275:1–275:46, 2021. URL https://jmlr.org/papers/v22/20-445.html.

Garud N. Iyengar. Robust dynamic programming. *Math. Oper. Res.*, 30(2):257–280, 2005. doi: 10.1287/MOOR.1040.0129. URL https://doi.org/10.1287/moor.1040.0129.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artif. Intell.*, 101(1-2):99–134, 1998. doi: 10.1016/S0004-3702(98)00023-X. URL https://doi.org/10.1016/S0004-3702(98)00023-X.

David L. Kaufman and Andrew J. Schaefer. Robust modified policy iteration. *INFORMS J. Comput.*, 25(3):396–410, 2013. doi: 10.1287/IJOC.1120.0509. URL https://doi.org/10.1287/ijoc.1120.0509.

Hassan K Khalil and Jessy W Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

Merlijn Krale, Eline M. Bovy, Maris F. L. Galesloot, Thiago D. Simão, and Nils Jansen. On evaluating policies for robust POMDPs. In *Eighteenth European Workshop on Reinforcement Learning*, 2025. URL https://openreview.net/forum?id=Q1qM9iF3iH.

Hadas Kress-Gazit, Georgios E. Fainekos, and George J. Pappas. Temporal-logic-based reactive mission and motion planning. *IEEE Trans. Robotics*, 25(6):1370–1381, 2009. doi: 10.1109/TRO.2009.2030225. URL https://doi.org/10.1109/TRO.2009.2030225.

Tobias Meggendorfer, Maximilian Weininger, and Patrick 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*, pp. 26631–26641, 2025. doi: 10.1609/AAAI.V39I25.34865. URL https://doi.org/10.1609/aaai.v39i25.34865.

Hideaki Nakao, Ruiwei Jiang, and Siqian Shen. Distributionally robust partially observable markov decision process with moment-based ambiguity. *SIAM J. Optim.*, 31(1):461–488, 2021. doi: 10.1137/19M1268410. URL https://doi.org/10.1137/19M1268410.

Arnab Nilim and Laurent El 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]*, pp. 839–846, 2003. URL https://proceedings.neurips.cc/paper/2003/hash/300891a62162b960cf02ce3827bb363c-Abstract.html.

Takayuki 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*, pp. 106–115, 2015. URL http://proceedings.mlr.press/v37/osogami15.html.

Amir Pnueli. The temporal logic of programs. In *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*, pp. 46–57, 1977. doi: 10.1109/SFCS.1977.32. URL https://doi.org/10.1109/SFCS.1977.32.

Aurko Roy, Huan Xu, and Sebastian 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*, pp. 3043–3052, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/84c6494d30851c63a55cdb8cb047fadd-Abstract.html.

Koushik Sen, Mahesh Viswanathan, and Gul 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*, pp. 394–410, 2006. doi: 10.1007/11691372_26. URL https://doi.org/10.1007/11691372_26.

Marnix Suilen, Nils Jansen, Murat Cubuktepe, and Ufuk Topcu. Robust policy synthesis for uncertain pomdps via convex optimization. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pp. 4113–4120, 2020. doi: 10.24963/IJCAI.2020/569. URL https://doi.org/10.24963/ijcai.2020/569.

Marnix Suilen, Thom Badings, Eline M. Bovy, David Parker, and Nils 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*, pp. 126–154, 2024. doi: 10.1007/978-3-031-75778-5_7. URL https://doi.org/10.1007/978-3-031-75778-5_7.

Sean Summers and John Lygeros. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Autom.*, 46(12):1951–1961, 2010. doi: 10.1016/J.AUTOMATICA.2010.08.006. URL https://doi.org/10.1016/j.automatica.2010.08.006.

Aviv Tamar, Shie Mannor, and Huan 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*, pp. 181–189, 2014. URL http://proceedings.mlr.press/v32/tamar14.html.

Chen Tessler, Yonathan Efroni, and Shie 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*, pp. 6215–6224, 2019. URL http://proceedings.mlr.press/v97/tessler19a.html.

Ambuj Tewari and Peter 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*, pp. 263–277, 2007. doi: 10.1007/978-3-540-72927-3_20. URL https://doi.org/10.1007/978-3-540-72927-3_20.

Wolfgang Thomas. Languages, automata, and logic. In *Handbook of Formal Languages, Volume 3: Beyond Words*, pp. 389–455. 1997. doi: 10.1007/978-3-642-59126-6_7. URL https://doi.org/10.1007/978-3-642-59126-6_7.

Yue Wang and Shaofeng 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*, pp. 7193–7206, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/3a4496776767aaa99f9804d0905fe584-Abstract.html.

Yue Wang and Shaofeng Zou. Policy gradient method for robust reinforcement learning. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, pp. 23484–23526, 2022. URL https://proceedings.mlr.press/v162/wang22at.html.

Yue Wang, Alvaro Velasquez, George K. Atia, Ashley Prater-Bennette, and Shaofeng 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*, pp. 15215–15223, 2023. doi: 10.1609/AAAI.V37I12.26775. URL https://doi.org/10.1609/aaai.v37i12.26775.

Yue Wang, Alvaro Velasquez, George K. Atia, Ashley Prater-Bennette, and Shaofeng Zou. Robust average-reward reinforcement learning. *J. Artif. Intell. Res.*, 80:719–803, 2024. doi: 10.1613/JAIR.1.15451. URL https://doi.org/10.1613/jair.1.15451.

Wolfram Wiesemann, Daniel Kuhn, and Berç Rustem. Robust markov decision processes. *Math. Oper. Res.*, 38(1):153–183, 2013. doi: 10.1287/MOOR.1120.0566. URL https://doi.org/10.1287/moor.1120.0566.

Zhuodong Yu, Ling Dai, Shaohang Xu, Siyang Gao, and Chin Pang Ho. Fast bellman updates for wasserstein distributionally robust mdps. 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. URL http://papers.nips.cc/paper_files/paper/2023/hash/61779e9b0c26a31c5f36bd3e8c180dcf-Abstract-Conference.html.

Appendix

This is the technical appendix of the submitted paper "Solving Robust POMDPs with ω-regular Objectives via Partially Observable Stochastic Games"

Table of Contents

- Section A Prescription compilation for ac-POSGs
- Section B Conversion of RPOMDP $\mathcal{M}$ to POSG $\mathcal{G}$
  - Section B.1 Mapping of plays, observation histories and strategies from $\mathcal{M}$ to $\mathcal{G}$
  - Section B.2 Lifting objectives from $\mathcal{M}$ to $\mathcal{G}$
  - Section B.3 Value Preservation between $\mathcal{M}$ to $\mathcal{G}$
  - Section B.4 Proof of Theorem 3.1
- Section C ac-POSG $\mathcal{G}$ to Pre-min transformed POSG $\mathcal{G}'$
  - Section C.1 Mapping of plays, observation histories and strategies from $\mathcal{G}$ to $\mathcal{G}'$
  - Section C.2 Lifting objectives from $\mathcal{G}$ to $\mathcal{G}'$
  - Section C.3 Value Preservation between $\mathcal{G}$ to $\mathcal{G}'$
  - Section C.4 Proof of Lemma 3.2
- Section D Constructing an RPOMDP from a Pre-min-transformed POSG
  - Section D.1 Mapping of plays, observation histories and strategies from $\mathcal{G}'$ to $\mathfrak{R}(\mathcal{G})$
  - Section D.2 Lifting objectives from $\mathcal{G}'$ to $\mathfrak{R}(\mathcal{G})$
  - Section D.3 Value Preservation between $\mathcal{G}'$ to $\mathfrak{R}(\mathcal{G})$
  - Section D.4 Proof of Lemma 3.3

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.

Henceforth, all the strategies in the below proofs are *effective strategies*: they are defined only on effective actions.

A Prescription Compilation for ac-POSGs

For $o \in \mathcal{O}_{\mathsf{min}}$, let observation class $S_{\mathsf{min}}(o) = \{t \in S_{\mathsf{min}} : \mathcal{Z}_{\mathsf{min}}(t) = o\}$ and fix an enumeration $S_{\mathsf{min}}(o) = \{t^o_1, \dots, t^o_{k_o}\}$. Let $k = \max_o k_o \geq 1$ and write $j(t)$ for the index of t in its observation class i.e., $t = t^{\mathcal{Z}_{\mathsf{min}}(t)}_{j(t)} \in S_{\mathsf{min}}(\mathcal{Z}_{\mathsf{min}}(t))$. The nonempty effective actions are denoted by $A_{\mathsf{max}}(s)$ and $A_{\mathsf{min}}(t)$ for $s \in S_{\mathsf{max}}, t \in S_{\mathsf{min}}$; they equal the global alphabet when no restriction is specified.

To compile factored prescriptions into ordinary ones, we introduce a gadget which can be plugged in at each min state. These gadgets have k selection stages, even when $|S_{\mathsf{min}}(o)| < k$. The common length prevents the max-player from learning the min-player's observation class from the duration of a gadget.

Given an ac-POSG $\mathcal{G}$ in which the min-player uses factored prescriptions rather than ordinary strategies, we compile $\mathcal{G}$ into an ac-POSG $\mathcal{G}^c$ whose min-player strategies are ordinary, in the sense Chatterjee et al. (2013). Intuitively, $\mathcal{G}^c$ simulates a single prescription choice at a min-state $t \in S_{\mathsf{min}}$ by unfolding it into k sequential selection stages, one per min-state, so an ordinary strategy can supply the k local actions one at a time sequentially, instead of all at once, as a prescription does. Between consecutive selection stages, a forced max move (with a fresh action $\top$) preserves strict alternation between the two players without making any real decision. Once all k stages have been visited, the accumulated choices are handed off through an exit state $D(t, \beta)$, via a fixed reference action b_0, into the transition $\delta(t, \beta)$ of the original state t that the assembled prescription selected. The construction below makes this precise.

Definition A.1 (Compiled ac-POSG)**.** Given an ac-POSG $\mathcal{G} = (S_{\mathsf{max}}, S_{\mathsf{min}}, A_{\mathsf{max}}, A_{\mathsf{min}}, \delta, s_0, \mathcal{O}_{\mathsf{max}}, \mathcal{O}_{\mathsf{min}}, \mathcal{Z}_{\mathsf{max}}, \mathcal{Z}_{\mathsf{min}})$ with priority function $p^{\mathcal{G}}$, define the ac-POSG $\mathcal{G}^c = (S^c_{\mathsf{max}}, S^c_{\mathsf{min}}, A^c_{\mathsf{max}}, A^c_{\mathsf{min}}, \delta^c, s_0, \mathcal{O}^c_{\mathsf{max}}, \mathcal{O}^c_{\mathsf{min}}, \mathcal{Z}^c_{\mathsf{max}}, \mathcal{Z}^c_{\mathsf{min}})$ as follows.

- *States.*Define

$$U_{t,i} = \begin{cases} \{\bot\} & i \leq j(t), \\ A_{\mathsf{min}}(t) & i > j(t), \end{cases} \qquad V_{t,i} = \begin{cases} \{\bot\} & i < j(t), \\ A_{\mathsf{min}}(t) & i \geq j(t). \end{cases}$$

  where $\bot$ is a fresh symbol. The selection states are $S^{\mathrm{sel}}_{\mathsf{min}} = \{(t, i, \beta) : t \in S_{\mathsf{min}},\ 1 \leq i \leq k,\ \beta \in U_{t,i}\}$, and the exit states are $E = \{D(t, \beta) : t \in S_{\mathsf{min}},\ \beta \in A_{\mathsf{min}}(t)\}$. Set $S^c_{\mathsf{min}} = S^{\mathrm{sel}}_{\mathsf{min}} \cup E$ and $S^c_{\mathsf{max}} = S_{\mathsf{max}} \cup M$ where $M = \{F(t, i, \beta) : t \in S_{\mathsf{min}},\ 1 \leq i \leq k,\ \beta \in V_{t,i}\}$. $\beta = \bot$ before stage $j(t)$ and contains an effective action afterwards.

- *Actions.* Choose distinct fresh actions $b_0, \top \notin A_{\mathsf{max}} \cup A_{\mathsf{min}}$ and set

$$A^c_{\mathsf{min}} = A_{\mathsf{min}} \cup \{b_0\}, \qquad A^c_{\mathsf{max}} = A_{\mathsf{max}} \cup \{\top\}.$$

  S_{max} retain $A_{\mathsf{max}}(s)$, and max-states in M have sole effective action $\top$. At a selection state of class o, the effective actions are given by

$$B_{o,i} = \begin{cases} A_{\mathsf{min}}(t^o_i) & 1 \leq i \leq k_o, \\ \{b_0\} & k_o < i \leq k. \end{cases}$$

  Every exit state has sole effective action b_0. The fresh action never belongs to a source min-prescription and is never recorded in the register β.

- *Transitions.* Write $t = t^o_j$ where $o = \mathcal{Z}_{\mathsf{min}}(t)$ and $1 \le j \le k_o$. For $s \in S_{\mathsf{max}}, a \in A_{\mathsf{max}}(s)$: $\delta^c(s,a)((t,1,\bot)) := \delta(s,a)(t)$. For $(t,i,\beta) \in S^{\mathrm{sel}}_{\mathsf{min}}$ and $b \in B_{o,i}$,

$$\delta^c\big((t,i,\beta),b\big) := \mathrm{Dirac}(F(t,i,\beta')), \qquad \beta' = \begin{cases} b & \text{if } i = j \\ \beta & \text{if } i \neq j. \end{cases}$$

For $F(t,i,\beta) \in S^c_{\mathsf{max}}$ with $i < k$, $\delta^c\big(F(t,i,\beta),\top\big) := \mathrm{Dirac}((t,i+1,\beta))$; for the state $F(t,k,\beta)$ (where $\beta \in A_{\mathsf{min}}(t)$),

$$\delta^c\big(F(t,k,\beta),\top\big) := \mathrm{Dirac}(D(t,\beta)).$$

For the exit state $D(t,\beta)$, the fresh action b_0 is the only effective action, and $\delta^c\big(D(t,\beta),b_0\big) := \delta(t,\beta)$.

- *Observations.* For $\star \in \{\mathsf{max},\mathsf{min}\}$:

$$\mathcal{O}^c_\star = \mathcal{O}_\star \uplus \{(o,i) : o \in O_\star,\ 1 \le i \le k+1\}.$$

Define $\mathcal{Z}^c_\star(s) = \mathcal{Z}_\star(s)$ for $s \in S_{\mathsf{max}}$ and

$$\mathcal{Z}^c_\star((t,i,\beta)) = \mathcal{Z}^c_\star(F(t,i,\beta)) = (\mathcal{Z}_\star(t),i), \qquad \mathcal{Z}^c_\star(D(t,\beta)) = (\mathcal{Z}_\star(t),k+1).$$

Thus each player retains its original observation of t and learns only the public stage index, not t, β, or the other player's observation.

- *Priorities.* $p^c(s) = p^{\mathcal{G}}(s)$ for $s \in S_{\mathsf{max}}$, and $p^c\big((t,i,\beta)\big) = p^c\big(F(t,i,\beta)\big) = p^c\big(D(t,\beta)\big) := p^{\mathcal{G}}(t)$.

Only the transitions listed above are effective. Every effective transition alternates owners.

Every min-state t is unfolded into a chain of k min steps, alternating with max-states in M carrying the dummy action $\top$. At stage i, the min-player declares what it would play if the true state were t^o_i; only the declaration made at the stage matching the true state is recorded (in β). After the k_o component choices, $k - k_o$ forced padding choices use b_0. The exit $D(t,\beta)$ then executes $\delta(t,\beta)$ on b_0. The fresh min-action thus carries no additional decision. Both compiled players use standard observation-based action distributions, with the effective-action support convention stated below. Figure 3 illustrates the gadget for a class of size two, padded to $k = 3$.

Lemma A.2 (Size and transitions of $\mathcal{G}^c$)**.** *Put $n = |S_{\mathsf{max}}|$, $m = |S_{\mathsf{min}}|$, $a = |A_{\mathsf{max}}|$, $b = |A_{\mathsf{min}}|$, and $\ell_t = |A_{\mathsf{min}}(t)|$ for $t \in S_{\mathsf{min}}$. Let $r_1 = \sum_{s \in S_{\mathsf{max}}} |A_{\mathsf{max}}(s)|$ and $r_2 = \sum_{t \in S_{\mathsf{min}}} \ell_t$ count the effective state–action rows in $\mathcal{G}$, and let e_1, e_2 count their positive-probability edges in $\mathcal{G}$.*

For $t \in S_{\mathsf{min}}$ with stage index $j = j(t)$: a selection state (t,i,β) still has $\beta = \bot$ only, before β is recorded ($i \le j$, one choice); from stage j on, β ranges over $A_{\mathsf{min}}(t)$ ($i > j$, ℓ_t choices). The forced state $F(t,i,\beta)$ that follows it in the same stage is one step ahead: it sees $\beta = \bot$ only strictly before stage j ($i < j$), and the recorded value already from stage j on ($i \ge j$). Summing over all k stages gives the exact state counts

$$C = \sum_{t \in S_{\mathsf{min}}} \big(j(t) + (k - j(t))\ell_t\big) = |S^{\mathrm{sel}}_{\mathsf{min}}|, \qquad H = \sum_{t \in S_{\mathsf{min}}} \big(j(t) - 1 + (k - j(t) + 1)\ell_t\big) = |M|.$$

Each selection state has one outgoing transition per effective action, giving the total

$$R = \sum_{t \in S_{\mathsf{min}}} \sum_{i=1}^{k} |U_{t,i}|\,|B_{\mathcal{Z}_{\mathsf{min}}(t),i}|$$

of transitions leaving selection states.

For the construction in Definition A.1, the exact counts are

$$\begin{aligned} |S^c_{\mathsf{min}}| &= C + r_2, & |S^c_{\mathsf{max}}| &= n + H, \\ |A^c_{\mathsf{min}}| &= b + 1, & |A^c_{\mathsf{max}}| &= a + 1, \\ r^c &= r_1 + R + H + r_2, & e^c &= e_1 + R + H + e_2, \\ |\mathcal{O}^c_\star| &= (k+2)|O_\star| \qquad (\star \in \{\mathsf{max},\mathsf{min}\}). \end{aligned}$$

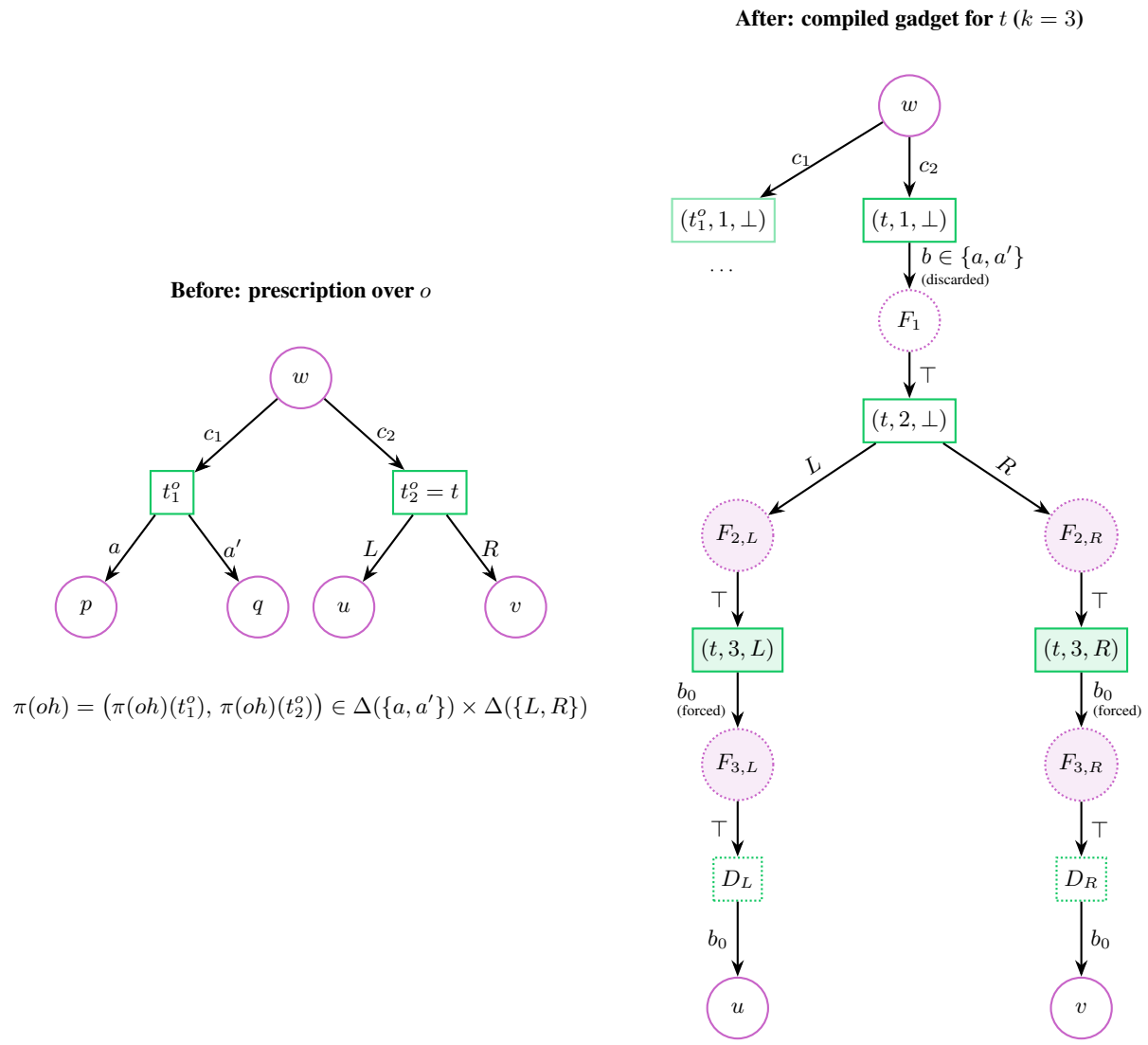

Figure 3: The compiled gadget for $t = t_2^o$ when $S_{\mathsf{min}}(o) = \{t_1^o, t_2^o\}$, so $k_o = 2$, drawn with $k = 3$: elsewhere in the same game, a different observation class o' has $k_{o'} = 3$ states, and since every gadget is padded to the global $k = \max_{o''} k_{o''}$, this class's gadget must also run for 3 stages even though it has only 2 states of its own. On the left, both states of o are reachable from a common max-state w, via distinct actions c_1 and c_2 each leading deterministically to t_1^o and t_2^o respectively. On the right, w's transitions are unchanged: c_1 and c_2 still lead to t_1^o and t_2^o, but each now enters their respective gadgets rather than the original state directly – $(t_1^o, 1, \bot)$ is shown only as a stub. Following the c_2 branch, the three stages are consumed one at a time: Stage 1 (state $F_1 = F(t, 1, \bot)$) chooses $b \in \{a, a'\}$, discarded on the displayed branch since the play never visits t_1^o; Stage 2 records $b \in \{L, R\}$, reaching $F_{2,L} = F(t, 2, L)$ or $F_{2,R} = F(t, 2, R)$; Stage 3 has no state of o left to represent, so it is a forced padding stage – the only effective action there is b_0 – reaching $F_{3,L} = F(t, 3, L)$ or $F_{3,R} = F(t, 3, R)$, and finally the exits $D_L = D(t, L)$ or $D_R = D(t, R)$. The fresh min-action $b_0 \notin \{L, R\}$ is forced at both the padding stage and either exit. States differing only in the stored action have identical observations.

Here r^c counts effective transition distributions and e^c counts positive triples (s, a, s'), not merely state pairs. In particular,

$$|S^c| = O(n + km(b+1)), \qquad r^c = O(r_1 + kmb^2), \qquad e^c = O(e_1 + e_2 + kmb^2).$$

The construction is polynomial in the explicit input size. When $k = 1$, its size is linear in the source size; alternatively the identity game can be used because a min-prescription already has a single component.

Proof. The state counts $C = |S^{\mathrm{sel}}_{\mathsf{min}}|$ and $H = |M|$ follow directly from the case split above; adding the r_2 exit states (one per t and $\beta \in A_{\mathsf{min}}(t)$) gives $|S^c_{\mathsf{min}}| = C + r_2$, and $|S^c_{\mathsf{max}}| = n + H$.

For transitions: the rows inherited unchanged from $\mathcal{G}$, $\delta^c(s,a)((t,1,\bot)) := \delta(s,a)(t)$ for $s \in S_{\mathsf{max}}, a \in A_{\mathsf{max}}(s)$, number exactly r_1 with e_1 edges. Selection states contribute the R transitions counted above. Every forced state has exactly one outgoing transition, giving H rows and edges. Each exit state $D(t,\beta)$ has the single effective action b_0 and copies the source min-transition $\delta(t,\beta)$ exactly, giving r_2 rows and e_2 edges. Summing gives r^c and e^c.

The observation count follows from the disjoint tagged alphabets: $\mathcal{O}^c_\star = \mathcal{O}_\star \uplus \{(o,i) : o \in O_\star, 1 \le i \le k+1\}$ has $|O_\star| + (k+1)|O_\star| = (k+2)|O_\star|$ elements.

Since $|U_{t,i}|, |V_{t,i}|, |B_{o,i}| \le b$, the polynomial bounds follow by bounding each sum by its number of terms times b (or b^2 for R). Probabilities are copied or equal to 1, with no increase in rational bit length. For $k = 1$, every t has $j(t) = 1$, so $C = m$, $H = R = r_2$, and the formulas give a linear-size expansion. □

Lemma A.3 (Prescription Compilation)**.** *Let $\mathcal{G}$ be an ac-POSG in which the max-player uses ordinary observation-based strategies and the min-player uses factored prescriptions, and let $\mathcal{G}^c$ be its compiled game (Definition A.1). Then $\mathcal{G}^c$ is an ac-POSG in which both players use ordinary observation-based strategies. Each player's strategies admit mutually inverse maps $\Phi_\star, \Psi_\star$ ($\star \in \{\mathsf{max}, \mathsf{min}\}$), such that for every source pair (σ,π) and compiled pair (σ^c,π^c),*

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

where W is a parity objective and W^c is its lift (priorities as in Definition A.1). Moreover, a play can happen under (σ,π) in $\mathcal{G}$ if and only if a corresponding play can happen under the lifted strategies in $\mathcal{G}^c$. Sure winning needs exactly this: it depends on which plays are possible, not on their probabilities. Hence sure, almost-sure, limit-sure, and quantitative analysis of W in $\mathcal{G}$ reduce in polynomial time to the same analyses of W^c in $\mathcal{G}^c$, and back, at the construction size of Lemma A.2. Memoryless, finite-memory, and infinite-memory strategies correspond exactly, with no increase in the number of memory states. If the min-player already has full observation ($k = 1$), $\mathcal{G} = \mathcal{G}^c$.

Proof. Objective preservation. Every visit to t stretches into $2k+1$ gadget states, all carrying priority $p^{\mathcal{G}}(t)$ by construction. A parity objective depends only on which priorities occur infinitely often along a play. If ρ^c is obtained from ρ by locally repeating each visited priority some bounded number of times (here at most $2k+1$), the set of priorities occurring infinitely often is unchanged: a priority visited infinitely often in ρ is still visited infinitely often in ρ^c, since each visit contributes at least one copy, and a priority visited only finitely often in ρ still contributes only finitely many copies in ρ^c, since each of its finitely many visits contributes at most $2k+1$ copies. So a play ρ satisfies W in $\mathcal{G}$ exactly when its counterpart ρ^c satisfies W^c in $\mathcal{G}^c$.

Plays. A single visit to a min-state $t = t^o_j$ in a play of $\mathcal{G}$ unfolds, in $\mathcal{G}^c$, into a full gadget traversal: k selection stages interleaved with forced max-states, then the exit $D(t,\beta)$. Only the action played at stage j matters; every other stage's action is discarded and can be anything. So one $\mathcal{G}$-play corresponds to many $\mathcal{G}^c$-plays, and there is no useful function between them at the level of full plays. We work with observation histories instead, where the correspondence is clean.

Observation histories. Every visited state, min-owned or max-owned, contributes an observation to each player's history (Section 3), so one gadget contributes *two* tagged observations per stage: one from the selection state (t,i,β), one from the forced state $F(t,i,\beta)$ right after it, both tagged $(\mathcal{Z}_\star(t),i)$ by construction. Write

$$\mathcal{B}(o) = (o,1),(o,1),(o,2),(o,2),\ldots,(o,k),(o,k),(o,k+1)$$

for a complete gadget's tag sequence, ending in the lone $(o,k+1)$ since the exit $D(t,\beta)$ has no forced state after it.

A min-observation history in $\mathcal{G}^c$ is a sequence of complete blocks $\mathcal{B}(o)$ for finished gadgets, followed optionally by a partial block recording the stage reached in the current one. Let $\Theta_{\mathsf{min}}(oh,i)$ build this

partial history for a source history oh ending at $o = \mathrm{last}(oh)$: expand every earlier observation to a complete block, and expand o itself only up to stage i ($1 \le i \le k+1$; $i = k+1$ means the block is complete). Let Υ_{min} collapse any block, complete or partial, back to the single observation it belongs to. These are mutual inverses,

$$\Upsilon_{\mathsf{min}}(\Theta_{\mathsf{min}}(oh, i)) = oh, \qquad \Theta_{\mathsf{min}}(\Upsilon_{\mathsf{min}}(oh^c), i) = oh^c, \qquad \text{where } \mathrm{last}(oh^c) = (o, i) \text{ for some } o \in \mathcal{O}_{\mathsf{min}}.$$

For the first identity: $\Theta_{\mathsf{min}}(oh, i)$ is, by construction, one block per observation of oh, in the same order – complete blocks for every observation but the last, a partial block for the last – so collapsing each block back to the single observation it belongs to just reads off oh again. For the second: every reachable min-observation history of $\mathcal{G}^c$ really does have this shape, since a gadget is entered only at its first tag $(o, 1)$ and left only once its exit tag $(o, k{+}1)$ is reached, so no history can end mid-gadget in any other way; consequently $\Upsilon_{\mathsf{min}}(oh^c)$ recovers the same observation sequence oh that produced oh^c, and the stage i is exactly the length of oh^c's trailing partial block, i.e. $\mathrm{last}(oh^c) = (o, i)$. Re-expanding every observation of oh but the last into a complete block, and the last into its first i tags, therefore reconstructs oh^c tag for tag. The max-player's maps $\Theta_{\mathsf{max}}, \Upsilon_{\mathsf{max}}$ are defined the same way, over max-observation histories.

Strategies. At a history oh^c ending at stage i for observation o, define the strategy mappings $\Phi_{\mathsf{min}} : \Pi_{\mathcal{G}} \to \Pi_{\mathcal{G}^c}, \Psi_{\mathsf{min}} : \Pi_{\mathcal{G}^c} \to \Pi_{\mathcal{G}}$ as follows:

$$\Phi_{\mathsf{min}}(\pi)(oh^c) := \pi(\Upsilon_{\mathsf{min}}(oh^c))(t_i^o) \ (i \le k_o), \qquad \Phi_{\mathsf{min}}(\pi)(oh^c) := \mathrm{Dirac}(b_0) \ (i > k_o, \text{ or at an exit state}),$$
$$\Psi_{\mathsf{min}}(\pi^c)(oh)(t_i^o) := \pi^c(\Theta_{\mathsf{min}}(oh, i)).$$

Informally, $\Phi_{\mathsf{min}}(\pi)$ answers each stage of a gadget the way the source prescription would have answered at the state that stage stands for: at stage $i \le k_o$, it plays whatever π would play at t_i^o; once i exceeds k_o, or at the exit, there is no state left to represent, so it simply plays b_0. Conversely, $\Psi_{\mathsf{min}}(\pi^c)$ recovers π's i-th component by asking the compiled strategy π^c what it plays once told, via $\Theta_{\mathsf{min}}(oh, i)$, that stage i of the fully-expanded gadget for t_i^o's observation has been reached. And symmetrically $\Phi_{\mathsf{max}}(\sigma)$ plays $\sigma(\Upsilon_{\mathsf{max}}(oh^c))$ at an S_{max}-state and $\top$ at a forced state, with $\Psi_{\mathsf{max}}(\sigma^c)(oh) := \sigma^c(\Theta_{\mathsf{max}}(oh))$. Each right-hand side is supported on the correct effective actions: $A_{\mathsf{min}}(t_i^o)$ at a selection stage, $\{b_0\}$ at a stage past k_o or at the exit, $\{\top\}$ at a forced state.

These maps are mutually inverse because $\Theta_\star, \Upsilon_\star$ are. For a selection-stage history $oh^c = \Theta_{\mathsf{min}}(oh, i)$ with $i \le k_o$,

$$\Psi_{\mathsf{min}}(\Phi_{\mathsf{min}}(\pi))(oh)(t_i^o) = \Phi_{\mathsf{min}}(\pi)(\Theta_{\mathsf{min}}(oh, i)) = \pi(\Upsilon_{\mathsf{min}}(\Theta_{\mathsf{min}}(oh, i)))(t_i^o) = \pi(oh)(t_i^o),$$

so $\Psi_{\mathsf{min}} \circ \Phi_{\mathsf{min}} = \mathrm{id}$; substituting the other way, using $\Theta_{\mathsf{min}}(\Upsilon_{\mathsf{min}}(oh^c), i) = oh^c$, gives $\Phi_{\mathsf{min}} \circ \Psi_{\mathsf{min}} = \mathrm{id}$. Stages past k_o and the max-player's maps invert the same way.

Probability preservation. Fix a history ending at $\min$-state $t = t_j^o$. Under (σ, π) in $\mathcal{G}$, the next action is $b \sim \pi(oh)(t)$ and the next max-state is drawn from $\delta(t, b)$.

Under $(\Phi_{\mathsf{max}}(\sigma), \Phi_{\mathsf{min}}(\pi))$ in $\mathcal{G}^c$: at every stage $i \ne j$, whichever action is played, $\delta^c((t, i, \beta), b) = \mathrm{Dirac}(F(t, i, \beta'))$ lands at the same place regardless of b (since $\beta' = \beta$ when $i \ne j$), so the total continuation probability is $\sum_b \Phi_{\mathsf{min}}(\pi)(oh_i)(b) = 1$ no matter what is played. At stage j, $\Phi_{\mathsf{min}}(\pi)$ plays exactly $\pi(oh)(t)$, recording $\beta = b \sim \pi(oh)(t)$. Every forced state passes through deterministically, and the exit plays b_0 deterministically, so

$$\Pr_{\mathcal{G}^c}[\text{exit state} = s'] = \sum_b \pi(oh)(t)(b)\, \delta(t, b)(s') = \Pr_{\mathcal{G}}[\text{next state} = s'].$$

By induction on history length, corresponding cylinder sets under (σ, π) and $(\Phi_{\mathsf{max}}(\sigma), \Phi_{\mathsf{min}}(\pi))$ have equal probability, giving the first displayed equality for any measurable W; the reverse equality follows the same way using $\Psi_{\mathsf{max}}, \Psi_{\mathsf{min}}$.

The same idea gives support preservation, in plain terms: a $\mathcal{G}^c$-play matches a possible $\mathcal{G}$-play exactly when the action recorded at stage j is one the strategy could really play there. Whatever is played at every other stage doesn't matter, so long as it's also something the strategy could really play, since any such choice still projects down to the same $\mathcal{G}$-play. So nothing about which plays are possible in $\mathcal{G}$ is lost or added by moving to $\mathcal{G}^c$.

Together with probability and support preservation, this gives $\mathrm{Val}_{\mathcal{G}}(W) = \mathrm{Val}_{\mathcal{G}^c}(W^c)$, since Val is the supremum over max-strategies of the infimum over min-strategies of the matching probabilities,

and those probabilities agree pairwise under $\Phi_{\mathsf{max}}, \Phi_{\mathsf{min}}$ (and $\Psi_{\mathsf{max}}, \Psi_{\mathsf{min}}$). The same ingredients give the other three analyses directly:

- *Sure winning* holds for (σ, π) in $\mathcal{G}$ iff every play it can produce satisfies W, iff – by play correspondence and objective equivalence – every play $(\Phi_{\mathsf{max}}(\sigma), \Phi_{\mathsf{min}}(\pi))$ can produce in $\mathcal{G}^c$ satisfies W^c, iff the lifted pair is sure winning in $\mathcal{G}^c$.
- *Almost-sure winning* holds iff the matching probability equals 1, preserved directly by the displayed probability equality.
- *Limit-sure and quantitative winning* are threshold statements about $\mathrm{Val}_{\mathcal{G}}(W)$ and $\mathrm{Val}_{\mathcal{G}^c}(W^c)$, already shown equal above.

Memory regimes. A finite-memory strategy is realized by a transducer $T = (Q, q_0, u, g)$: Q is a finite memory set with initial state q_0, u updates the memory on each new observation, and $g(q)$ outputs the strategy's full action assignment at memory state q, so the strategy played after observation history oh is $g(q)$ for q the memory state reached by applying u along oh from q_0. The stage index i is never hidden – it is written into the observation (o, i) itself – so no extra memory is ever needed just to track it.

From π to $\Phi_{\mathsf{min}}(\pi)$. Build T^c over the same Q: hold memory fixed across a gadget, updating only once, on its closing tag: $u^c(q, (o, k{+}1)) := u(q, o)$, and $u^c(q, (o, i)) := q$ otherwise. Output $g^c(q, (o, i)) := g(q)(t_i^o)$ at a selection stage ($i \le k_o$) and $\mathrm{Dirac}(b_0)$ at a stage past k_o or at the exit.

From π^c to $\Psi_{\mathsf{min}}(\pi^c)$. Build T over the same Q: to process one source observation o, replay T^c over $\mathcal{B}(o)$ without storing intermediate states, $u(q, o) := u^{c,*}(q, \mathcal{B}(o))$, where $u^{c,*}(q, w)$ denotes u^c applied one tag at a time along the sequence w, starting from q (so $u^{c,*}(q, \varepsilon) = q$ and $u^{c,*}(q, w \cdot x) = u^c(u^{c,*}(q, w), x)$ for a tag x); to compute the output at stage i, stop the same replay just before (o, i), $g(q)(t_i^o) := g^c(u^{c,*}(q, \mathcal{B}_{<i}(o)), (o, i))$. Both computations reuse Q, with no extra memory.

That these transducers realize $\Phi_{\mathsf{min}}(\pi)$ and $\Psi_{\mathsf{min}}(\pi^c)$ follows from $\Theta_{\mathsf{min}}, \Upsilon_{\mathsf{min}}$ being mutual inverses, exactly as in the Strategies part. The max-player's transducers expand and contract the same way, with forced output $\mathrm{Dirac}(\top)$ at every forced state. Both directions therefore preserve the number of memory states exactly: memoryless, finite-memory, and infinite-memory strategies all carry over unchanged. When $k = 1$, $\mathcal{G}^c = \mathcal{G}$ and the transducers coincide. □

Example 1. *Let t be a min-state in an observation class $\{t_1^o, t_2^o\}$ of size $k_o = 2$, with $t = t_2^o$, and suppose the game's global stage count is $k = 3$, set by a different observation class elsewhere with 3 states. Take $A_{\mathsf{min}}(t_1^o) = \{a, a'\}$, $A_{\mathsf{min}}(t) = A_{\mathsf{min}}(t_2^o) = \{L, R\}$, $\delta(t, L) = \mathrm{Dirac}(u)$, and $\delta(t, R) = \mathrm{Dirac}(v)$ for max-states u, v. The compiled min-alphabet is $\{a, a', b_0\}$ at stage 1, $\{L, R, b_0\}$ at stage 2, and $\{b_0\}$, forced, at stage 3, with $b_0 \notin \{a, a', L, R\}$. Figure 3 shows the compiled gadget. A prescription value $\pi(oh)(t) = \{L \mapsto 0.3, R \mapsto 0.7\}$ is realized by playing $\pi(oh)(t_1^o) \in \{a, a'\}$ at stage 1 (discarded on the displayed t_2^o branch), $\{L \mapsto 0.3, R \mapsto 0.7\}$ at stage 2, the forced b_0 at the padding stage 3, and the fresh b_0 deterministically at the exit state $D(t, \beta)$; the exit distribution over $\{u, v\}$ is exactly $\{u \mapsto 0.3, v \mapsto 0.7\}$, matching $\mathcal{G}$. This branch has four selection states, five forced max-states, and two exits; the action stored in the two exits is L or R, never b_0.*

Lemma A.4 (Ordinary strategies as prescriptions)**.** *Let $\widehat{\Pi}_{\mathcal{G}}$ denote the set of ordinary observation-based min-strategies $\widehat{\pi} : \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}} \to \Delta(A_{\mathsf{min}})$, with no restriction on which actions $\widehat{\pi}$ supports at any history – in particular $\widehat{\pi}(oh)$ need not avoid actions ineffective at any state of $\mathrm{last}(oh)$'s class. The* broadcast map $\iota_{\mathsf{min}} : \widehat{\Pi}_{\mathcal{G}} \to \Pi_{\mathcal{G}}$, *defined by*

$$\iota_{\mathsf{min}}(\widehat{\pi})(oh)(t) := \widehat{\pi}(oh) \qquad \textit{for every } t \in S_{\mathsf{min}}(\mathrm{last}(oh)),$$

is a bijection from $\widehat{\Pi}_{\mathcal{G}}$ onto $\Pi_{\mathcal{G}}^0 := \iota_{\mathsf{min}}(\widehat{\Pi}_{\mathcal{G}})$, the class-constant *prescriptions – those $\pi \in \Pi_{\mathcal{G}}$ with $\pi(oh)(t) = \pi(oh)(t')$ for all $t, t' \in S_{\mathsf{min}}(\mathrm{last}(oh))$ – and it preserves outcome probabilities exactly:*

$$\mathbb{P}_{\mathcal{G}}^{\sigma, \widehat{\pi}}(W) = \mathbb{P}_{\mathcal{G}}^{\sigma, \iota_{\mathsf{min}}(\widehat{\pi})}(W) \qquad \textit{for every } \sigma \in \Sigma_{\mathcal{G}},\ \widehat{\pi} \in \widehat{\Pi}_{\mathcal{G}},\ \textit{and measurable } W. \qquad (*)$$

Consequently the game value, and the sure, almost-sure, limit-sure, and quantitative analyses of $\mathcal{G}$, are exactly the same whether the min-player ranges over $\widehat{\Pi}_{\mathcal{G}}$ or over $\Pi_{\mathcal{G}}^0$:

$$\mathrm{Val}_{\mathcal{G}}^{\widehat{\Pi}_{\mathcal{G}}}(W) \;=\; \mathrm{Val}_{\mathcal{G}}^{\Pi_{\mathcal{G}}^0}(W), \qquad \textit{where} \qquad \mathrm{Val}_{\mathcal{G}}^{X}(W) := \sup_{\sigma \in \Sigma_{\mathcal{G}}} \inf_{\pi \in X} \mathbb{P}_{\mathcal{G}}^{\sigma, \pi}(W).$$

A prescription outside $\Pi^0_{\mathcal{G}}$ genuinely discriminates within a class – min-player strategies are then strictly more expressive than $\widehat{\Pi}_{\mathcal{G}}$ – and it is exactly for such prescriptions that Lemma A.3's compiled game is needed; nothing here asserts $\mathrm{Val}^{\widehat{\Pi}_{\mathcal{G}}}_{\mathcal{G}}(W) = \mathrm{Val}^{\Pi_{\mathcal{G}}}_{\mathcal{G}}(W)$ for the full *prescription space, which can fail once prescriptions outside $\Pi^0_{\mathcal{G}}$ are allowed to compete.*

Proof. Well-defined. $\iota_{\mathsf{min}}(\widehat{\pi})(oh)$ assigns the single distribution $\widehat{\pi}(oh) \in \Delta(A_{\mathsf{min}})$ to every $t \in S_{\mathsf{min}}(o)$, $o = \mathrm{last}(oh)$, so it lies in $\prod_{t \in S_{\mathsf{min}}(o)} \Delta(A_{\mathsf{min}})$, as the prescription type requires.

Bijective onto $\Pi^0_{\mathcal{G}}$. For any $t \in S_{\mathsf{min}}(\mathrm{last}(oh))$, $\widehat{\pi}(oh) = \iota_{\mathsf{min}}(\widehat{\pi})(oh)(t)$, so $\widehat{\pi}$ is recovered from $\iota_{\mathsf{min}}(\widehat{\pi})$ by reading off any one component – this shows ι_{min} is injective. Every $\iota_{\mathsf{min}}(\widehat{\pi})$ is class-constant by construction, so $\Pi^0_{\mathcal{G}} \subseteq \{\text{class-constant prescriptions}\}$; conversely, given a class-constant π, setting $\widehat{\pi}(oh) := \pi(oh)(t)$ for any $t \in S_{\mathsf{min}}(\mathrm{last}(oh))$ is well-defined (class-constancy says the choice of t does not matter) and satisfies $\iota_{\mathsf{min}}(\widehat{\pi}) = \pi$. So ι_{min} is onto the class-constant prescriptions exactly, and, being injective, a bijection $\widehat{\Pi}_{\mathcal{G}} \to \Pi^0_{\mathcal{G}}$.

Probability preservation, $(*)$. At a min-state $t = t^o_j$, playing $\widehat{\pi}$ draws the next action from $\widehat{\pi}(oh)$; playing $\iota_{\mathsf{min}}(\widehat{\pi})$ draws it from $\iota_{\mathsf{min}}(\widehat{\pi})(oh)(t) = \widehat{\pi}(oh)$ – the same distribution, since $t \in S_{\mathsf{min}}(o)$. Every other component of the game – the states, σ's response, the transition function δ – is literally unchanged. So $(\sigma, \widehat{\pi})$ and $(\sigma, \iota_{\mathsf{min}}(\widehat{\pi}))$ induce identical transition probabilities at every step of every history; by induction on history length and 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, giving $(*)$ for any measurable W, with no size blow-up and no change to memory requirements.

Value preservation. Fix $\sigma \in \Sigma_{\mathcal{G}}$. By $(*)$, the bijection $\widehat{\pi} \mapsto \iota_{\mathsf{min}}(\widehat{\pi})$ from $\widehat{\Pi}_{\mathcal{G}}$ to $\Pi^0_{\mathcal{G}}$ matches each ordinary strategy with a prescription achieving the *same* probability of W against σ. Hence the two sets of achievable probabilities coincide exactly, not just their bounds:

$$\left\{\mathbb{P}^{\sigma,\widehat{\pi}}_{\mathcal{G}}(W) : \widehat{\pi} \in \widehat{\Pi}_{\mathcal{G}}\right\} = \left\{\mathbb{P}^{\sigma,\iota_{\mathsf{min}}(\widehat{\pi})}_{\mathcal{G}}(W) : \widehat{\pi} \in \widehat{\Pi}_{\mathcal{G}}\right\} = \left\{\mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W) : \pi \in \Pi^0_{\mathcal{G}}\right\},$$

the last equality because ι_{min} is onto $\Pi^0_{\mathcal{G}}$. Two sets of reals that coincide have equal infima, so

$$\inf_{\widehat{\pi} \in \widehat{\Pi}_{\mathcal{G}}} \mathbb{P}^{\sigma,\widehat{\pi}}_{\mathcal{G}}(W) = \inf_{\pi \in \Pi^0_{\mathcal{G}}} \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W) \qquad \text{for every } \sigma \in \Sigma_{\mathcal{G}}.$$

This holds pointwise in σ; taking $\sup_{\sigma \in \Sigma_{\mathcal{G}}}$ of both sides – valid since the two sides agree termwise, with no further structure on $\Sigma_{\mathcal{G}}$ needed – gives $\mathrm{Val}^{\widehat{\Pi}_{\mathcal{G}}}_{\mathcal{G}}(W) = \mathrm{Val}^{\Pi^0_{\mathcal{G}}}_{\mathcal{G}}(W)$. The same termwise argument, applied to $\{\sigma : \mathbb{P}^{\sigma,\widehat{\pi}}_{\mathcal{G}}(W) = 1 \ \forall \widehat{\pi}\}$ and its prescription counterpart (and likewise at any limit-sure or quantitative threshold), transfers sure, almost-sure, and limit-sure winning exactly between the two strategy spaces, since these notions are themselves defined from the same probabilities compared in $(*)$.

Theargument uses only that ι_{min} is an exact, probability-preserving bijection onto $\Pi^0_{\mathcal{G}}$. It does *not* extend to $\mathrm{Val}^{\Pi_{\mathcal{G}}}_{\mathcal{G}}(W)$ for the full prescription space: once prescriptions outside $\Pi^0_{\mathcal{G}}$ are allowed to compete in the infimum, the min-player may do strictly better by discriminating within a class, and $\mathrm{Val}^{\Pi_{\mathcal{G}}}_{\mathcal{G}}(W) \leq \mathrm{Val}^{\widehat{\Pi}_{\mathcal{G}}}_{\mathcal{G}}(W)$ can be strict as shown in example 2. □

Example 2 (The expressiveness gap is strict)**.** *Let $\mathcal{G}$ have max-states $S_{\mathsf{max}} = \{s_0, \mathsf{good}, \mathsf{bad}\}$, min-states t_1, t_2 sharing a single observation o, so $S_{\mathsf{min}}(o) = \{t_1, t_2\} = \{t^o_1, t^o_2\}$ ($k_o = 2$), and actions $A_{\mathsf{min}} = \{L, R\}$; s_0 has a single available action a. Transitions:*

$$\delta(s_0, a) := \tfrac{1}{2}\,\mathrm{Dirac}(t_1) + \tfrac{1}{2}\,\mathrm{Dirac}(t_2), \delta(t_1, L) := \mathrm{Dirac}(\mathsf{good}), \ \ \delta(t_1, R) := \mathrm{Dirac}(\mathsf{bad}),$$
$$\delta(t_2, L) := \mathrm{Dirac}(\mathsf{bad}), \ \ \delta(t_2, R) := \mathrm{Dirac}(\mathsf{good}).$$

good *and* bad *are absorbing with* good *max-winning and* bad *min-winning; take W to be reachability to* good *(e.g. priority 0 on* good*'s region, priority 1 elsewhere). Figure 4 shows both regimes. Since s_0 has only one available action, $\Sigma_{\mathcal{G}}$ is a singleton and $\mathrm{Val}^{X}_{\mathcal{G}}(W) = \inf_{\pi \in X} \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W)$ for the unique σ, whichever strategy space X ranges over. Both t_1, t_2 produce the identical min-observation history oh after s_0, a, since $\mathcal{Z}_{\mathsf{min}}(t_1) = \mathcal{Z}_{\mathsf{min}}(t_2) = o$.*

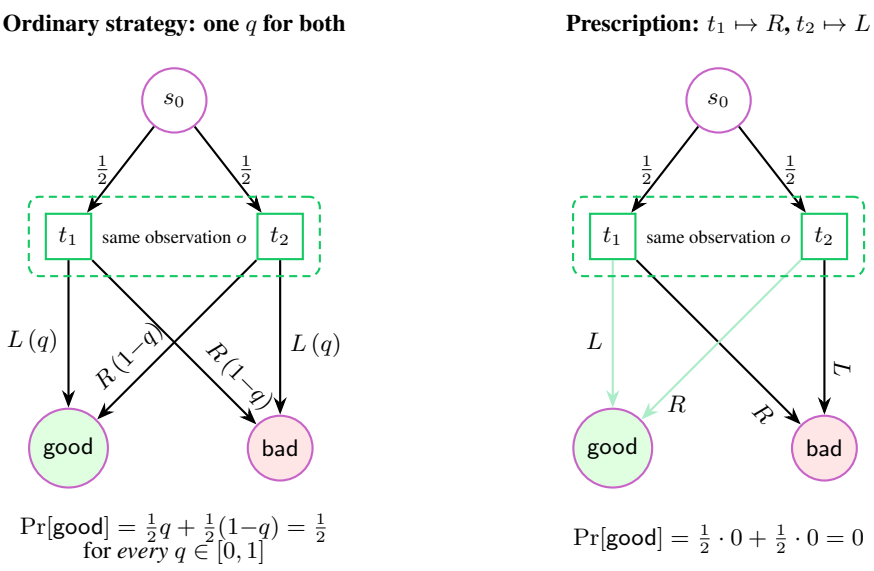

Figure 4: The observation class $S_{\mathsf{min}}(o) = \{t_1, t_2\}$ is dashed to mark that both states produce the same history oh, so any *ordinary* strategy (left) must answer them with the same distribution $(q, 1-q)$ over (L, R) – and every choice of q yields exactly $\Pr[\text{good}] = \frac{1}{2}$, since t_1, t_2 want opposite answers and the two errors always average out. A *prescription* (right) answers each state correctly and forces $\Pr[\text{good}] = 0$, faded edges marking the actions it never plays.

Ordinary strategies. $\widehat{\pi}(oh) = (q, 1-q)$ *over* (L, R)*, for any fixed* $q \in [0,1]$*, gives*

$$\mathbb{P}_{\mathcal{G}}^{\sigma,\widehat{\pi}}(W) = \tfrac{1}{2}\big(q \cdot 1 + (1-q)\cdot 0\big) + \tfrac{1}{2}\big(q \cdot 0 + (1-q)\cdot 1\big) = \tfrac{1}{2},$$

independently of q*: every ordinary strategy scores exactly* $\frac{1}{2}$*, so* $\mathrm{Val}_{\mathcal{G}}^{\widehat{\Pi}_{\mathcal{G}}}(W) = \frac{1}{2}$*. This is exactly the phenomenon underlying the well-formedness argument of Lemma A.4:* $\widehat{\pi}(oh)$ *must answer the same way regardless of which of* t_1, t_2 *is truly occupied, and here the two states want opposite answers, so no single* q *helps.*

Prescriptions. $\pi(oh)(t_1) := \mathrm{Dirac}(R)$*,* $\pi(oh)(t_2) := \mathrm{Dirac}(L)$ *sends both states to* bad *deterministically, so* $\mathbb{P}_{\mathcal{G}}^{\sigma,\pi}(W) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = 0$*, giving* $\mathrm{Val}_{\mathcal{G}}^{\Pi_{\mathcal{G}}}(W) = 0$*.*

B 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 $S^{\mathcal{G}}_{\mathsf{min}} = \{(s,a) : s \in S,\ a \in A_{\mathcal{M}}(s)\}$. For an original RPOMDP with every action effective, this is $S \times A$. Let $S^{\mathcal{G}} = S^{\mathcal{G}}_{\mathsf{max}} \cup S^{\mathcal{G}}_{\mathsf{min}}$.
- *Actions:* $A^{\mathcal{G}}_{\mathsf{max}} = A$ and $A^{\mathcal{G}}_{\mathsf{min}} = \bigcup_{(s,a)\in S^{\mathcal{G}}_{\mathsf{min}}} V_{s,a}$. The effective alphabets are $A^{\mathcal{G}}_{\mathsf{max}}(s) = A_{\mathcal{M}}(s)$ where $A_{\mathcal{M}}(s)$ denotes the set of effective actions at state s and $A^{\mathcal{G}}_{\mathsf{min}}((s,a)) = V_{s,a}$. Let $A^{\mathcal{G}} = A^{\mathcal{G}}_{\mathsf{max}} \cup A^{\mathcal{G}}_{\mathsf{min}}$.
- *Transition function* $\delta^{\mathcal{G}}$ on the effective state–action pairs is defined as follows:
  - For $s \in S^{\mathcal{G}}_{\mathsf{max}}$ and $a \in A_{\mathcal{M}}(s)$, the transition function is the Dirac distribution with probability mass on $(s,a) \in S^{\mathcal{G}}_{\mathsf{min}}$

$$\delta^{\mathcal{G}}(s,a)(s') = \begin{cases} 1 & \text{if } s' = (s,a) \\ 0 & \text{otherwise} \end{cases}$$

– 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

$$\delta^{\mathcal{G}}((s,a),v)(s') = \begin{cases} v(s') & \text{where } s' \in S^{\mathcal{G}}_{\mathsf{max}} \text{ and } v(s') \text{ denotes the } s'\text{th entry in } v \\ 0 & \text{otherwise} \end{cases}$$

Other min-actions have probability zero by the standing strategy convention.

- *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 $\mathcal{O}^{\mathcal{G}}_{\mathsf{max}} = \mathcal{O}_{\mathsf{a}}$ and $\mathcal{O}^{\mathcal{G}}_{\mathsf{min}} = \mathcal{O}_{\mathsf{e}}$ 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 $(s,a) \in S^{\mathcal{G}}_{\mathsf{min}}$, $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}((s,a)) = \mathcal{Z}_{\mathsf{e}}(s)$

A 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 $k \geq 0$, $s_{2k} \in S^{\mathcal{G}}_{\mathsf{max}}$, $a_{2k} \in A_{\mathcal{M}}(s_{2k})$, $s_{2k+1} = (s_{2k}, a_{2k}) \in S^{\mathcal{G}}_{\mathsf{min}}$, and $a_{2k+1} \in V_{s_{2k},a_{2k}}$. 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 $o^2_i = \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s_i)$.

B.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, $\forall k \geq 0 : s'_{2k} = s_k, a'_{2k} = a_k, s'_{2k+1} = (s'_{2k}, a'_{2k}) = (s_k, a_k), a'_{2k+1} = v \in \{v \in V_{s_k,a_k} \mid v(s_{k+1}) > 0\}$ (arbitrary v), since $P(s_k, a_k)(s_{k+1}) > 0$ and $P(s_k, a_k) \in \mathrm{conv}(V_{s_k,a_k})$, this subset is non-empty. . 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 $X \subseteq \mathsf{Plays}_{\mathcal{M}}$, $\quad \Theta(X) = \bigcup_{\rho \in X} \Theta(\rho)$.
- 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 B.1. *Let $\star \in \{a, e\}$. $\Theta_\star$ and $\Upsilon_\star$ are inverses of each other.*

Proof.

1. For all $oh = o_0, o_0, o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$,
$$\Theta_a(\Upsilon_a(oh)) = \Theta_a(o_0, o_1, o_2, \cdots, o_{t-1}, o_t) = o_0, o_0, o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t = oh$$
2. For all $oh = o_0, o_1, o_2, \cdots, o_{t-1}, o_t \in \mathsf{OH}_a$,
$$\Upsilon_a(\Theta_a(oh)) = \Upsilon_a(o_0, o_0, o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t) = o_0, o_1, o_2, \cdots, o_{t-1}, o_t = oh$$
3. For all $oh = o_0, o_0, o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t, o_t \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}}$,
$$\Theta_e(\Upsilon_e(oh)) = \Theta_e(o_0, o_1, o_2, \cdots, o_{t-1}, o_t) = o_0, o_0, o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t, o_t = oh$$
4. For all $oh = o_0, o_1, o_2, \cdots, o_{t-1}, o_t \in \mathsf{OH}_e$,
$$\Upsilon_e(\Theta_e(oh)) = \Upsilon_e(o_0, o_0, o_1, o_1, o_2, o_2, \cdots, o_{t-1}, o_{t-1}, o_t, o_t) = o_0, o_1, o_2, \cdots, o_{t-1}, o_t = oh$$

□

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

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

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
$$\sigma : \mathsf{OH}_a \to \Delta(A).$$
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
$$\sigma' : \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}} \to \Delta(A^{\mathcal{G}}_{\mathsf{max}}).$$
4. The min-player strategy $\pi' \in \Pi_{\mathcal{G}}$ for POSG $\mathcal{G}$ from the reduction is given by
$$\pi' : \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}} \to \bigcup_{o \in \mathcal{O}^{\mathcal{G}}_{\mathsf{min}}} \prod_{s \in S^{\mathcal{G}}_{\mathsf{min}}(o)} \Delta(A^{\mathcal{G}}_{\mathsf{min}})$$
Every compatible component $\pi'(oh)((s, a))$ is supported on $V_{s,a}$.

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.

$$\mathfrak{D}_e(\pi')(oh,a)(x,b) = \begin{cases} \mathsf{bar}_{x,a}(\pi'(\Theta_e(oh))((x,a))) & \text{if } b = a \in A_{\mathcal{M}}(x) \text{ and } \mathcal{Z}_{\mathsf{e}}(x) = \mathrm{last}(oh), \\ P^0_{x,b} & \text{otherwise.} \end{cases}$$

  Hence every component belongs to its corresponding local uncertainty set.

Remark B.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 B.3. *The agent maps $\mathfrak{C}_a$ and $\mathfrak{D}_a$ are mutual inverses. The environment maps satisfy, at every compatible oh and effective pair (s, a) with $\mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s) = \mathrm{last}(oh)$,*

$$\mathfrak{D}_e(\mathfrak{C}_e(\pi))(oh, a)(s, a) = \pi(oh, a)(s, a)$$
$$\mathsf{bar}_{s,a}(\mathfrak{C}_e(\mathfrak{D}_e(\pi'))(oh)((s, a))) = \mathsf{bar}_{s,a}(\pi'(oh)((s, a)))$$

Thus both compositions preserve the next-state distribution.

Proof. The agent identities follow from Lemma B.1: both maps compose observation histories and copy a single action distribution. The local effective max alphabets agree with the RPOMDP agent alphabets, so the maps preserve support. For every min-strategy, $\pi'(oh)((s, a)) \in \Delta(V_{s,a})$; hence the reverse barycenter is well-defined without a further strategy restriction. The environment identities follow from $\mathsf{bar}_{s,a}(\alpha^P_{s,a}) = P$ and the two history identities in Lemma B.1.

$$\mathfrak{D}_e(\mathfrak{C}_e(\pi))(oh, a)(s, a) = \mathsf{bar}_{s,a}(\mathfrak{C}_e(\pi)(\Theta_e(oh))((s, a))) = \pi(\Upsilon_e(\Theta_e(oh)), a)(s, a) = \pi(oh, a)(s, a)$$
$$\mathsf{bar}_{s,a}(\mathfrak{C}_e(\mathfrak{D}_e(\pi'))(oh)((s, a))) = \mathfrak{D}_e(\pi')(\Upsilon_e(oh), a)(s, a) = \mathsf{bar}_{s,a}(\pi'(\Theta_e(\Upsilon_e(oh)))((s, a))) = \mathsf{bar}_{s,a}(\pi'(oh)((s, a)))$$

□

B.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 by 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 B.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 $\rho' = s_0, a_0, (s_0, a_0), a'_1, s_1, \cdots$ 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. □

B.3 Value Preservation

Lemma B.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 cylinder set ($\mathrm{Cyl}(h)$) 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 $h = s_0, (o_0^a, o_0^e), s_1, \dots s_t, (o_t^a, o_t^e) \in H_{\mathcal{M}}$. Then $\Theta_h(h)$ is the set of all histories $h' = s'_0, (o_0^{max}, o_0^{min}), s'_1, \dots s'_{2t}, (o_{2t}^{max}, o_{2t}^{min})$ 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 $\forall i \geq 0$ $o_i^{max} = \mathcal{Z}^{\mathcal{G}}_{\mathsf{max}}(s'_i), o_i^{min} = \mathcal{Z}^{\mathcal{G}}_{\mathsf{min}}(s'_i)$. Thus, $\Theta_h(h)$ consists of all histories in G obtained by resolving each transition in h using an arbitrary choice of action from $A^{\mathcal{G}}_{\mathsf{min}}$ at each step.

We show that for a finite history h of every play in W, there exists corresponding set of finite histories $H' = \Theta_h(h)$ of some plays 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 observation 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 observation histories for history h_1. Let oh^{max} be the 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 $h'_1 \in H'_1$. 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 $\sigma'(oh^{\mathsf{max}})(a)$ 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., $\left(\sum_{v \in V_{s,a}} \pi'(oh^{\mathsf{min}})((s, a))(v) \cdot v(s_1)\right)$.

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

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

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

The converse cylinder identity follows by the same induction, replacing the vertex distribution by its barycenter. The second identity of Lemma B.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 B.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 B.5 gives $q_{\mathcal{G}}(\mathfrak{C}_a(\sigma)) \le q_{\mathcal{M}}(\sigma)$, because every $\pi \in \Pi_{\mathcal{M}}$ has the image $\mathfrak{C}_e(\pi) \in \Pi_{\mathcal{G}}$.

$$q_{\mathcal{M}}(\sigma) = \inf_{\pi \in \Pi_{\mathcal{M}}} \mathbb{P}_{\mathcal{M}}^{\sigma,\pi}(W) = \inf_{\mathfrak{C}_e(\pi) \in \Pi_{\mathcal{G}}} \mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\mathfrak{C}_e(\pi)}(W') \ge \inf_{\pi' \in \Pi_{\mathcal{G}}} \mathbb{P}_{\mathcal{G}}^{\mathfrak{C}_a(\sigma),\pi'}(W') = q_{\mathcal{G}}(\mathfrak{C}_a(\sigma))$$

Conversely, for every $\pi' \in \Pi_{\mathcal{G}}$, the second equality in Lemma B.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

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

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 B.3, so taking the outer supremum yields

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

□

Lemma B.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

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

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 B.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 B.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 B.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 B.7,

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

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 B.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 B.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,*

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

Proof. The fixed-strategy equality established in the proof of Lemma B.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 B.10 (Limit-sure Analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and $W' = \Theta(W)$ be the corresponding objective in $\mathcal{G}$. Then,*

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

Proof. Follows directly from Lemma B.6. □

Lemma B.11 (Quantitative Analysis)**.** *Let $W \subseteq \mathsf{Plays}_{\mathcal{M}}$ be a parity objective of $\mathcal{M}$ and $W' = \Theta(W)$ be 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 B.6. □

B.4 Proof of Theorem 3.1

Proof. Because parity subsumes all ω-regular objectives, from Lemma B.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 B.8, B.9, B.10, and B.11. □

C 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 C.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. The effective alphabets are $A'_{\mathsf{max}}(s) = A_{\mathsf{max}}(s)$ at original max-states, $A'_{\mathsf{max}}(d_s) = \{\top\}$, and $A'_{\mathsf{min}}(s) = A_{\mathsf{min}}(s)$ at original min-states. Every strategy uses these alphabets by definition.

- *Transition function δ':*
  - For $s_1 \in S_{\mathsf{max}}$, $a_1 \in A_{\mathsf{max}}(s_1)$, 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 $a_2 \in A_{\mathsf{min}}(s_2)$: $\delta'(s_2, a_2) = \delta(s_2, a_2)$.
  - For $d_s \in D$ and action $\top$: $\delta'(d_s, \top) = \mathrm{Dirac}(s)$.

- *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 $\mathcal{Z}'_{\mathsf{min}}(d_s) = o^{\#}_{\mathcal{Z}_{\mathsf{min}}(s)}$.

C.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 C.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 C.3. *Every play $h' = s_0, a_0, s_1, a_1, s_2, a_2, \cdots \in \mathsf{Plays}_{\mathcal{G}'}$ satisfies, for each $k \geq 0$:*

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

Lemma C.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 C.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 by removing all occurrences of observations in $\mathcal{O}^{\#}$.

Lemma C.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 C.7. *For all $oh' = (oh'_1, o) \in \mathsf{OH}^{\mathcal{G}'}_{\star}$,*

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

Mapping of strategies. As in the main text, max-strategies output a single action distribution, whereas min-strategies output a prescription:

$$\sigma : \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}} \to \Delta(A_{\mathsf{max}}), \qquad \pi : \mathsf{OH}^{\mathcal{G}}_{\mathsf{min}} \to \bigcup_{o \in \mathcal{O}_{\mathsf{min}}} \prod_{s \in S_{\mathsf{min}}(o)} \Delta(A_{\mathsf{min}}).$$

The corresponding types in $\mathcal{G}'$ use A'_{max} and A'_{min}. All distributions obey the effective-action support convention. In particular every max-strategy in $\mathcal{G}'$ plays $\top$ at a dummy observation, not only strategies produced by the lifting map.

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 $\sigma \in \Sigma_{\mathcal{G}}$ and $oh' \in \mathsf{OH}^{\mathcal{G}'}_{\mathsf{max}}$, define
$$\Phi_{\mathsf{max}}(\sigma)(oh') = \begin{cases} \sigma((\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh')) & \mathrm{last}(oh') \notin \mathcal{O}^{\dagger}, \\ \mathrm{Dirac}(\top) & \mathrm{last}(oh') \in \mathcal{O}^{\dagger}. \end{cases}$$
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 $\sigma' \in \Sigma_{\mathcal{G}'}$ and $oh \in \mathsf{OH}^{\mathcal{G}}_{\mathsf{max}}$, define
$$\Psi_{\mathsf{max}}(\sigma')(oh) = \sigma'(\Gamma^{\mathsf{max}}_{\mathcal{O}}(oh)).$$
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 C.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. For max, Lemma C.6 gives, as an equality of action distributions,

$$\Psi_{\mathsf{max}}(\Phi_{\mathsf{max}}(\sigma))(oh) = \Phi_{\mathsf{max}}(\sigma)(\Gamma^{\mathsf{max}}_{\mathcal{O}}(oh)) = \sigma((\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1}(\Gamma^{\mathsf{max}}_{\mathcal{O}}(oh))) = \sigma(oh).$$

Conversely, at a history oh' ending at an original max-state, Lemma C.7 gives

$$\Phi_{\mathsf{max}}(\Psi_{\mathsf{max}}(\sigma'))(oh') = \Psi_{\mathsf{max}}(\sigma')((\Gamma_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh')) = \sigma'(\Gamma_{\mathcal{O}}^{\mathsf{max}}((\Gamma_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh'))) = \sigma'(oh').$$

At a dummy history, both sides equal $\mathrm{Dirac}(\top)$ by the strategy definition. The copied distributions have the same effective alphabet at the corresponding original observation, so both maps are well-defined on all max-strategies.

For min, the maps preserve the local action alphabet at each original min-state. We first show $\Psi_{\mathsf{min}}(\Phi_{\mathsf{min}}(\pi))(oh)(s) = \pi(oh)(s)$ for all $\pi \in \Pi_{\mathcal{G}}$, $oh \in \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}}$ and $s \in S_{\mathsf{min}}$ such that $\mathcal{Z}_{\mathsf{min}}(s) = \mathrm{last}(oh)$.

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

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

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

□

C.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 C.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. □

C.3 Value Preservation

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

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

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 of every play 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

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

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 $s_1 = \mathrm{last}(h'_1)$.

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

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 $a \in A_{\mathsf{min}}(s_1)$: By definition of Φ_{min} and $oh_1 = (\Gamma^{\mathsf{min}}_{\mathcal{O}})^{-1}(oh'_1)$ 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 original effective min state–action pairs, we have $\delta'(s_1,a)(s) = \delta(s_1,a)(s)$. Hence

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

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 $s_1 = \mathrm{last}(h'_1) = \mathrm{last}(h_1)$.

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

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 $a \in A_{\mathsf{max}}(s_1)$ and $oh_1 = (\Gamma^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh'_1)$, the strategy definition gives $\Phi_{\mathsf{max}}(\sigma)(oh'_1)(a) = \sigma(oh_1)(a)$. By construction, $\delta'(s_1,a)(d_s) = \delta(s_1,a)(s)$. Since $\top$ is the sole effective action at any dummy state, $\Phi_{\mathsf{max}}(\sigma)(oh'_1, o^{\dagger}_o)(\top) = 1$. Finally, $\delta'(d_s,\top)(s) = 1$ by construction of $\mathcal{G}'$. Hence

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

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 C.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 C.10 gives $q_{\mathcal{G}'}(\Phi_{\mathsf{max}}(\sigma)) \le q_{\mathcal{G}}(\sigma)$ because every $\pi \in \Pi_{\mathcal{G}}$ has an image $\Phi_{\mathsf{min}}(\pi) \in \Pi_{\mathcal{G}'}$.

$$q_{\mathcal{G}}(\sigma) = \inf_{\pi \in \Pi_{\mathcal{G}}} \mathbb{P}^{\sigma,\pi}_{\mathcal{G}}(W) = \inf_{\Phi_{\mathsf{min}}(\pi) \in \Pi_{\mathcal{G}'}} \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\Phi_{\mathsf{min}}(\pi)}_{\mathcal{G}'}(W') \ge \inf_{\pi' \in \Pi_{\mathcal{G}'}} \mathbb{P}^{\Phi_{\mathsf{max}}(\sigma),\pi'}_{\mathcal{G}'}(W') = q_{\mathcal{G}'}(\Phi_{\mathsf{max}}(\sigma))$$

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

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

Hence $q_{\mathcal{G}}(\sigma) = q_{\mathcal{G}'}(\Phi_{\mathsf{max}}(\sigma))$. Since Φ_{max} is a bijection by Lemma C.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 C.12. *Let $\mathcal{G}$ be any alternating control POSG and $\mathcal{G}'$ its Pre-*min *transformation. Let p be a priority function for $\mathcal{G}$. Then*

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

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

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

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

where $\mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}(resp.\mathsf{Plays}_{\mathcal{G}'}^{\sigma',\pi'})$ 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_k is the max-player observation history of ρ at s_k. By definition of Φ_{max}:

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

  Because $\rho \in \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}$, action a_k is in the support of $\sigma(oh_k)$, hence also of $\Phi_{\mathsf{max}}(\sigma)(oh'_k)$. The 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 source play.

- 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}:

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

  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 $C.5$), 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 C.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,*

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

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 C.9, for any $\rho \in \mathsf{Plays}_{\mathcal{G}}^{\sigma,\pi}, \rho \in W \iff \Gamma(\rho) \in W'$. 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 C.13 and C.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 C.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,*

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

Proof. The fixed-strategy equality established in the proof of Lemma C.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 C.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 C.11. □

Lemma C.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 C.11. □

C.3.1 Size and Memory Preservation of the Reduction

Lemma C.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 displayed effective transitions, excluding the optional completion. 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 for $s \in S_{\mathsf{min}}$, $a \in A_{\mathsf{min}}(s)$, and for $s \in S_{\mathsf{max}}$, $a \in A_{\mathsf{max}}(s)$ 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 $oh = o_0, o_1, \ldots, o_{2n} \in \mathsf{OH}^{\mathcal{G}}_{\star}$, so that $|oh| = 2n + 1$. By construction,

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

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 before every $\min$-player original observation. Hence

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

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

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

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

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

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).

We prove that Φ_{max} preserves memory. The arguments for Φ_{min}, Ψ_{max}, Ψ_{min} are symmetric: replace $\mathcal{O}^{\dagger}_{\mathsf{max}}$ by $\mathcal{O}^{\#}_{\mathsf{min}}$ for the min versions, and swap $\Gamma^{\star}_{O}$ with $(\Gamma^{\star}_{O})^{-1}$ for the Ψ versions.

Let $M^{\mathcal{G}}_{\mathsf{max}}, M^{\mathcal{G}}_{\mathsf{min}}$ denote the mealy machine for $\mathsf{max}, \mathsf{min}$-player strategies in $\mathcal{G}$. Then, for $M^{\mathcal{G}}_{\mathsf{max}} = \{Q \cup \{q_0\}, q_0, \mathcal{O}_{\mathsf{max}} \cup (\mathcal{O}_{\mathsf{max}} \times \mathcal{O}_{\mathsf{max}}), \tau, \lambda\}$: $Q \cup q_0$ are finite set of memory states where q_0 is initial state. The transition function τ is defined as:

$$\begin{aligned} \tau : q_0 \times \mathcal{O}_{\mathsf{max}} &\to Q \\ Q \times \mathcal{O}_{\mathsf{max}} \times \mathcal{O}_{\mathsf{max}} &\to Q \end{aligned}$$

The output function λ is defined as:

$$\begin{aligned} \lambda : q_0 \times \mathcal{O}_{\mathsf{max}} &\to \Delta(A_{\mathsf{max}}) \\ Q \times \mathcal{O}_{\mathsf{max}} \times \mathcal{O}_{\mathsf{max}} &\to \Delta(A_{\mathsf{max}}) \end{aligned}$$

Let $\sigma \in \Sigma_{\mathcal{G}}$ be realized by a Mealy machine $M^{\mathcal{G}}_{\mathsf{max}} = \{Q \cup q_0, q_0, \mathcal{O}_{\mathsf{max}} \cup \mathcal{O}_{\mathsf{max}} \times \mathcal{O}_{\mathsf{max}}, \tau, \lambda\}$ as defined above. We build a mealy machine $M^{\mathcal{G}'}_{\mathsf{max}} = \{Q' \cup q'_0 \cup Q_d, q'_0, \mathcal{O}'_{\mathsf{max}} \cup \mathcal{O}'_{\mathsf{max}} \times \mathcal{O}'_{\mathsf{max}}, \tau', \lambda'\}$ representing $\Phi_{\mathsf{max}}(\sigma)$ where $Q' = Q, q'_0 = q_0, Q_d = \{q_a \mid a \in Q\}$, for transition $\tau(q_0, o) = q_1$ converts to transition $\tau'(q_0, o) = q_{q_1}$ and $\tau(q_1, o_1, o_2) = q_2$ converts to 2 transitions $\tau'(q_{q_1}, o_{o_1}) = q_1, \tau'(q_1, o_1, o_2) = q_{q_2}$. The output λ, λ' are mapped as follows: $\lambda'(q_0, o) = \lambda(q_0, o)$ and $\lambda(q_1, o_1, o_2) = q_2$ gives two outputs as $\lambda'(q_{q_1}, o_{o_1})$ gives dirac distribution over $\top$, $\lambda'(q_1, o_1, o_2) = \lambda(q_1, o_1, o_2)$.

The mealy machine $M^{\mathcal{G}'}_{\mathsf{max}}$ realizes $\Phi_{\mathsf{max}}(\sigma)$ because it simulates the behavior of $M^{\mathcal{G}}_{\mathsf{max}}$ on observation histories after applying the inverse mapping $(\Gamma_O^{\mathsf{max}})^{-1}$. By construction, every transition in $M^{\mathcal{G}}_{\mathsf{max}}$ is expanded in $M^{\mathcal{G}'}_{\mathsf{max}}$ wherever required by introducing copied states Q_d that correspond to the play structure induced by Γ. The copied states Q_d handle the $\mathcal{O}^\dagger$ observations while producing $\mathrm{Dirac}(\top)$ when required. Because every observation history oh' in $\mathcal{G}'$ is reduced via $(\Gamma_O^{\mathsf{max}})^{-1}$ before applying the original strategy, the output function λ' mirrors λ exactly on the corresponding histories of G. Hence, the constructed machine produces the same action distributions as σ, establishing that it correctly realizes $\Phi_{\mathsf{max}}(\sigma)$.

Special cases. A memoryless strategy corresponds to $|Q| = 1$, which the construction trivially preserves. An infinite-memory strategy places no bound on $|Q|$; the construction uses the same Q, so the class is preserved.

□

C.4 Proof of Lemma 3.2

Proof. Because parity objective subsumes all ω-regular objectives, from lemma C.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 C.14, C.15, C.16, and C.17. □

D 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 C.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 each round visits an original max-state, its sampled dummy successor, and the corresponding min-state, in that order. The key observation is that the min-player's choice of action $a \in A_{\mathsf{min}}(s)$ 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 D.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 effective alphabets $A_{\mathfrak{R}}(s) = A_{\mathsf{max}}(s)$ for $s \in S_{\mathsf{max}}$ and $A_{\mathfrak{R}}(d_s) = \{\top\}$. Every agent strategy is one distribution over $A_{\mathfrak{R}}$, supported on the effective alphabet of its current observation class.
- *(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 $s \in S_{\mathsf{max}}$, $a \in A_{\mathsf{max}}(s)$, or $s = d_{s'} \in D$, $a = \top$, define

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

    For these pairs,

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

    where $\mathrm{conv}(V_{s,a})$ denotes the convex hull (polytope) of $V_{s,a}$. Remaining, ineffective pairs may use fixed singleton distributions; their components are never evaluated by an agent strategy. The list $V_{s,a}$ may contain redundant convex-hull generators. No claim that distinct min-actions induce distinct extreme points is needed.
- *(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 D.2. *The generating set $V_{d_s,\top} = \{\delta'(s,a) \mid a \in A_{\mathsf{min}}(s)\}$ of the polytope $P(d_s,\top)$ encodes all effective* min*-actions at $s \in S_{\mathsf{min}}$: each listed distribution is induced by an effective min-action. The environment selects a convex combination of these distributions, exactly matching a randomized min-action at 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.*

D.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 $k \geq 0$:

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

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 $oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_m \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}}$ ($\tilde{o}_{2k+1} \in \mathcal{O}_{\mathsf{max}}, \tilde{o}_{2k+2} \in \mathcal{O}^{\dagger}$ for $k \geq 0$), by $(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh^{\mathfrak{R}}) = oh = o_1, o_2, \ldots, o_n$. Insert the reconstructed min-observation only after a dummy observation that is followed by another retained observation, not after a terminal dummy observation. For all indices within this length:

$$\begin{aligned} o_{3k+1} &= \tilde{o}_{2k+1}, \\ o_{3k+2} &= \tilde{o}_{2k+2}, \\ o_{3k+3} &= x \text{ such that } \tilde{o}_{2k+2} = o^{\dagger}_x \quad (2k+2 < m), \\ o_n &= \tilde{o}_m. \end{aligned}$$

where $n = 3k+1$ when $m = 2k+1$ and $n = 3k+2$ when $m = 2k+2$.

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 $\widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}$. Its inverse $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1} : \widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})} \to \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}'}$ is defined for $oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_m \in \widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}$ by $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh^{\mathfrak{R}}) = o_1, o_2, \ldots, o_{m+m/2}$ 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 D.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 $oh' \in \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'}$, $(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(\Lambda_{\mathcal{O}}^{\mathsf{max}}(oh')) = oh'$.*

Proof. Let

$$oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_m \in \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})}.$$

We consider the two possible forms of m.

If $m = 2k + 1$, then by definition of $(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}$,

$$o_{3i+1} = \tilde{o}_{2i+1}, \qquad o_{3i+2} = \tilde{o}_{2i+2} \qquad o_{3i+3} = x \text{ where } \tilde{o}_{2i+2} = o_x^{\dagger}$$

for all $0 \leq i < k$, and $o_{3k+1} = \tilde{o}_{2k+1}$. Hence, applying $\Lambda_{\mathcal{O}}^{\mathsf{max}}$ removes precisely the inserted observations o_{3i+3} and retains

$$\Lambda_{\mathcal{O}}^{\mathsf{max}}\big((\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh^{\mathfrak{R}})\big) = \Lambda_{\mathcal{O}}^{\mathsf{max}}(o_1, o_2, \cdots, o_{3k+1}) = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_{2k+1} = oh^{\mathfrak{R}}.$$

If $m = 2k + 2$, then by definition of $(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}$,

$$o_{3i+1} = \tilde{o}_{2i+1}, \qquad o_{3i+2} = \tilde{o}_{2i+2}, \qquad o_{3i+3} = x \text{ where } \tilde{o}_{2i+2} = o_x^{\dagger}$$

for $0 \leq i < k$. Also,

$$o_{3k+1} = \tilde{o}_{2k+1}, \qquad o_{3k+2} = \tilde{o}_{2k+2}.$$

Therefore, applying $\Lambda_{\mathcal{O}}^{\mathsf{max}}$ removes precisely the inserted observations o_{3i+3} and retains

$$o_{3i+1} = \tilde{o}_{2i+1}, \qquad o_{3i+2} = \tilde{o}_{2i+2}.$$

Thus

$$\Lambda_{\mathcal{O}}^{\mathsf{max}}\big((\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh^{\mathfrak{R}})\big) = \Lambda_{\mathcal{O}}^{\mathsf{max}}(o_1, o_2, \cdots, o_{3k+1}, o_{3k+2}) = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_{2k+2} = oh^{\mathfrak{R}}.$$

Conversely, let

$$oh' = o_1, o_2, \ldots, o_n \in \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'}.$$

By the definition of $\Lambda_{\mathcal{O}}^{\mathsf{max}}$, its image is obtained by retaining

$$o_{3k+1}, o_{3k+2}$$

and deleting every o_{3k+3}. Applying $(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}$ then reconstructs each deleted observation o_{3k+3} from $o_{o_{3k+3}}^{\dagger}$ whenever a subsequent retained observation exists. If the history ends at $3k + 1$, no observation is inserted after the terminal observation; if it ends at $3k + 2$, the reconstructed history ends at $3k + 2$. Hence every observation of oh' is restored at its original position, and

$$(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(\Lambda_{\mathcal{O}}^{\mathsf{max}}(oh')) = o_1, o_2, \ldots, o_n = oh'.$$

Therefore both compositions are the identity. □

Lemma D.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 $oh' \in \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}'}$, $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}\big(\Lambda_{\mathcal{O}}^{\mathsf{min}}(oh')\big) = oh'$.*

Proof. Let

$$oh^{\mathfrak{R}} = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_m \in \widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}.$$

Since $oh^{\mathfrak{R}}$ ends in $\mathcal{O}^{\#}$, m is even. Write $m = 2k + 2$ for some $k \geq 0$. By definition of $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}$,

$$o_{3i+1} = \tilde{o}_{2i+1}, \qquad o_{3i+2} = \tilde{o}_{2i+2}, \qquad o_{3i+3} = x \text{ where } \tilde{o}_{2i+2} = o_x^{\#},$$

for all $0 \leq i \leq k$. Hence, applying $\Lambda_{\mathcal{O}}^{\mathsf{min}}$ removes exactly the inserted observations o_{3i+3} and retains

$$o_{3i+1} = \tilde{o}_{2i+1}, \qquad o_{3i+2} = \tilde{o}_{2i+2}.$$

Consequently,

$$\Lambda_{\mathcal{O}}^{\mathsf{min}}\big((\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh^{\mathfrak{R}})\big) = \Lambda_{\mathcal{O}}^{\mathsf{min}}(o_1, o_2, \cdots, o_{3k+3}) = \tilde{o}_1, \tilde{o}_2, \ldots, \tilde{o}_{2k+2} = oh^{\mathfrak{R}}.$$

Conversely, let

$$oh' = o_1, o_2, \ldots, o_n \in \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}'}.$$

Since oh' ends in an observation in $\mathcal{O}_{\mathsf{min}}$, its length is of the form $n = 3k + 3$ for some $k \geq 0$. By definition of $\Lambda_{\mathcal{O}}^{\mathsf{min}}$,

$$\tilde{o}_{2i+1} = o_{3i+1}, \qquad \tilde{o}_{2i+2} = o_{3i+2},$$

for all $0 \leq i \leq k$. Moreover,

$$\tilde{o}_{2i+2} = o_{o_{3i+3}}^{\#},$$

so $(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}$ reconstructs o_{3i+3} as the unique x satisfying $\tilde{o}_{2i+2} = o_x^{\#}$. Thus every deleted observation o_{3i+3} is reconstructed at its original position, while the retained observations are restored by

$$o_{3i+1} = \tilde{o}_{2i+1}, \qquad o_{3i+2} = \tilde{o}_{2i+2}.$$

In particular, the final observation is reconstructed as

$$o_{3k+3} = x \quad \text{such that} \quad \tilde{o}_{2k+2} = o_x^{\#}.$$

Therefore,

$$(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}\big(\Lambda_{\mathcal{O}}^{\mathsf{min}}(oh')\big) = (\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(\tilde{o}_1, \tilde{o}_2, \cdots, \tilde{o}_{2k+2}) = o_1, o_2, \ldots, o_n = oh'.$$

Hence both compositions are the identity. □

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

- A max-strategy in $\mathcal{G}'$ has type $\sigma' : \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'} \to \Delta(A'_{\mathsf{max}})$; a min-strategy retains its state-indexed prescription type.
- An RPOMDP agent strategy has type $\sigma^{\mathfrak{R}} : \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})} \to \Delta(A_{\mathfrak{R}})$.
- 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 D.1 is a convex combination of the vertices $V_{d_s,\top} = \{\delta'(s, a) \mid a \in A_{\mathsf{min}}(s)\}$. We write this as

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

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

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

For every $P \in P(d_s, \top)$, fix a coefficient vector $\beta_s^P \in \Delta(A_{\mathsf{min}}(s))$ with $\mathsf{bar}_s^{\mathcal{G}'}(\beta_s^P) = P$. Such vectors exist by the convex-hull definition and need not be unique; the fixed choice makes the reverse map well-defined. 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 $\sigma' \in \Sigma_{\mathcal{G}'}$ and $oh^{\mathfrak{R}} \in \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})}$,

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

At a dummy observation both sides equal $\mathrm{Dirac}(\top)$, since every max-strategy in $\mathcal{G}'$ uses the sole effective dummy action.

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

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

3. $\Xi_{\mathsf{max}} : \Sigma_{\mathfrak{R}(\mathcal{G})} \to \Sigma_{\mathcal{G}'}$: for each $\sigma^{\mathfrak{R}} \in \Sigma_{\mathfrak{R}(\mathcal{G})}$ and $oh' \in \mathsf{OH}_{\mathsf{max}}^{\mathcal{G}'}$,

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

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

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

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

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

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

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

Thus the environment compositions preserve transition distributions.

Proof. The agent identities follow from Lemma D.3 by composition, as equalities of single action distributions. The history maps retain the current observation and effective agent/max alphabet, so both maps preserve admissibility, including the forced dummy action. The environment maps preserve local support because $\beta_s^P \in \Delta(A_{\mathsf{min}}(s))$ and every barycenter belongs to $P(d_s, \top)$. The environment identities follow from $\mathsf{bar}_s^{\mathcal{G}'}(\beta_s^P) = P$ and Lemma D.4. At non-dummy states the uncertainty set is a singleton. □

D.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 the priority function for $\mathcal{G}'$.

Lemma D.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*

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

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. □

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

Lemma D.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 a 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' of a every play in W' ending at a state in $S_{\mathsf{max}} \cup D$, there exists a corresponding history $h^{\mathfrak{R}}$ in for some play in $W^{\mathfrak{R}}$ such that

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

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 two 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 $a \in A_{\mathsf{max}}(s_1)$, $P(s_1, a)$ is a singleton $\{\delta'(s_1, a)\}$. Then

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

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

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

Case 2: 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.

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 $\mathfrak{R}(\mathcal{G})$ for h'_1, s is $h_1^{\mathfrak{R}}$ (since S_{min}-state s is removed by Λ). Since the $\top$ action 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

$$\mathbb{P}^{\Omega_{\mathsf{max}}(\sigma'),\Omega_{\mathsf{min}}(\pi')}_{\mathfrak{R}(\mathcal{G})}(\mathrm{Cyl}\big(h_1^{\mathfrak{R}}\big)) = \mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(\mathrm{Cyl}(h'_1, s)) = \mathbb{P}^{\sigma',\pi'}_{\mathcal{G}'}(\mathrm{Cyl}(h'_1))$$

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.

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

By definition of Ω_{min},

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

Evaluating at s':

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

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

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

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 $a''\in A_{\mathsf{min}}(s)$ with probability $\pi'(\mathcal{Z}'_{\mathsf{min}}(h'_1,s))(s)(a'')$, reaching s' with probability $\delta'(s,a'')(s')$. Therefore, summing over all min actions in $\mathcal{G}'$:

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

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

For the converse mapping, repeat the cylinder induction with Ξ_{max} and Ξ_{min}. At a dummy state, the first environment identity in Lemma D.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 D.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 a 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

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

The first equality in Lemma D.8 gives $q_{\mathfrak{R}(\mathcal{G})}(\Omega_{\mathsf{max}}(\sigma')) \le q_{\mathcal{G}'}(\sigma')$.

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

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

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

so

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

Hence, $q_{\mathcal{G}'}(\sigma') = q_{\mathfrak{R}(\mathcal{G})}(\Omega_{\mathsf{max}}(\sigma'))$. Since Ω_{max} is a bijection by Lemma D.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 D.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*

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

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

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

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

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 $\Lambda_{\mathcal{O}}^{\mathsf{max}}$, and the definition of Ω_{max} gives

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

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 $\lambda = \pi'(oh'_t)(t) \in \Delta(A_{\mathsf{min}}(t))$. In $\mathfrak{R}(\mathcal{G})$, the mapped environment selects $\mathsf{bar}_t^{\mathcal{G}'}(\lambda)$ at $(d_t,\top)$. Hence, for every successor $s' \in S_{\mathsf{max}}$,

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

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 D.6 gives $\mathsf{bar}_t^{\mathcal{G}'}(\lambda) = \pi^{\mathfrak{R}}(\Lambda_{\mathcal{O}}^{\mathsf{min}}(oh'),\top)(d_t,\top)$, 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 D.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 D.11 and lemma D.7 gives

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

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

Lemma D.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 D.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 D.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 D.9. □

Lemma D.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 D.9. □

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

Lemma D.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 number of vertices in the uncertainity sets is at most $\sum_{t \in S_{\mathsf{min}}} |A_{\mathsf{min}}(t)| + \sum_{s \in S_{\mathsf{max}}} |A_{\mathsf{max}}(s)|$, 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 $oh^{\mathfrak{R}} \in \mathsf{OH}_{\mathsf{a}}^{\mathfrak{R}(\mathcal{G})}$, $|(\Lambda_{\mathcal{O}}^{\mathsf{max}})^{-1}(oh^{\mathfrak{R}})| \leq |oh^{\mathfrak{R}}| + \frac{|oh^{\mathfrak{R}}|-1}{2}$. For every $oh' \in \mathsf{OH}_{\mathsf{min}}^{\mathcal{G}'}$, $|\Lambda_{\mathcal{O}}^{\mathsf{min}}(oh')| = \frac{2}{3}|oh'|$ and for every $oh^{\mathfrak{R}} \in \widehat{\mathsf{OH}}_{\mathsf{e}}^{\mathfrak{R}(\mathcal{G})}$, $|(\Lambda_{\mathcal{O}}^{\mathsf{min}})^{-1}(oh^{\mathfrak{R}})| = \frac{3}{2}|oh^{\mathfrak{R}}|$*
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 $\mathcal{O}_{\mathsf{a}}^{\mathfrak{R}} = \mathcal{O}_{\mathsf{max}} \cup \{o_x^\dagger \mid x \in \mathcal{O}_{\mathsf{max}}\} = \mathcal{O}'_{\mathsf{max}}$ and $\mathcal{O}_{\mathsf{e}}^{\mathfrak{R}} = \mathcal{O}_{\mathsf{min}} \cup \{o_x^\# \mid x \in \mathcal{O}_{\mathsf{min}}\} = \mathcal{O}'_{\mathsf{min}}$, which have the same cardinalities as in $\mathcal{G}'$. For $s \in S_{\mathsf{max}}$ and $a \in A_{\mathsf{max}}(s)$, the vertex set $V_{s,a}$ is the singleton $\{\delta'(s,a)\}$. For each $(d_s, \top)$, it consists of $\delta'(s,a)$ for $a \in A_{\mathsf{min}}(s)$. Thus each source effective transition distribution is copied once, giving the stated sum of local alphabet sizes.

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

$$\Lambda_O^{\mathsf{max}}(oh') = o_0, z_1, \cdots, o_{2n},$$

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

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

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

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

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

$$|\Lambda_O^{\mathsf{max}}(oh')| = 2n+2 = \tfrac{2|oh'|+2}{3}.$$

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,

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

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

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

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,

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

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

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

Thus, for every $oh \in \mathsf{OH}^{\mathfrak{R}(\mathcal{G})}_{\mathsf{a}}$, $|(\Lambda^{\mathsf{max}}_{\mathcal{O}})^{-1}(oh)| \leq |oh| + \frac{|oh|-1}{2}$.

For environment history bound, $\Lambda^{\mathsf{min}}_{\mathcal{O}}$ removes a min-observation history for every 3 consecutive observations while its inverse adds min-observation for every 2 consective observations which creates the required bounds.

Clause (3).

We prove that Ω_{max} preserves memory. The arguments for Ω_{min}, Ξ_{max}, Ξ_{min} are symmetric: replace $\mathcal{O}^{\dagger}_{\mathsf{max}}$ by $\mathcal{O}^{\#}_{\mathsf{min}}$ for the min versions, and swap $\Lambda^{\star}_{O}$ with $(\Lambda^{\star}_{O})^{-1}$ for the Ξ versions and replace output of transitions by a prescriptions instead of single distribution.

Let $M^{\mathcal{G}'}_{\mathsf{max}}, M^{\mathcal{G}'}_{\mathsf{min}}$ denote the mealy machine for max, min-player strategies in $\mathcal{G}'$. Then, for $M^{\mathcal{G}'}_{\mathsf{max}} = \{Q \cup q_0 \cup Q_d, q_0, \mathcal{O}'_{\mathsf{max}} \cup \mathcal{O}'_{\mathsf{max}} \times \mathcal{O}'_{\mathsf{max}}, \tau', \lambda'\}$: $Q \cup q_0 \cup Q_d$ are finite set of memory states where q_0 is initial state and $Q_d = \{q_a \mid a \in Q\}$. The transition function τ' is defined as:

$$\begin{aligned} \tau' : q_0 \times \mathcal{O}_{\mathsf{max}} &\to Q_d \\ Q_d \times \mathcal{O}^{\dagger} &\to Q \\ Q \times \mathcal{O}_{\mathsf{max}} \times \mathcal{O}_{\mathsf{max}} &\to Q_d \end{aligned}$$

such that for $q_a \in Q_d, a \in Q, o \in \mathcal{O}^{\dagger}, \tau'(q_a, o) = a$

The output function λ' is defined as:

$$\begin{aligned} \lambda' : q_0 \times \mathcal{O}_{\mathsf{max}} &\to \Delta(A_{\mathsf{max}}) \\ Q_d \times \mathcal{O}^{\dagger} &\to \{\mathrm{Dirac}(\top)\} \\ Q \times \mathcal{O}_{\mathsf{max}} \times \mathcal{O}_{\mathsf{max}} &\to \Delta(A_{\mathsf{max}}) \end{aligned}$$

Let $\sigma' \in \Sigma_{\mathcal{G}'}$ be a strategy realized by mealy machine $M^{\mathcal{G}'}_{\mathsf{max}} = \{Q' \cup q'_0 \cup Q'_d, q'_0, \mathcal{O}'_{\mathsf{max}} \cup \mathcal{O}'_{\mathsf{max}} \times \mathcal{O}'_{\mathsf{max}}, \tau', \lambda'\}$ as defined above. We build a mealy machine $M^{\mathfrak{R}(\mathcal{G})}_{a} = (Q \cup q_0, q_0, \mathcal{O}'_{\mathsf{max}}, \tau, \lambda)$ representing $\Omega_{\mathsf{max}}(\sigma')$ where $Q = (Q' \times \mathcal{O}'_{\mathsf{max}}) \cup Q'_d, q_0 = q'_0$. The transition function is given by: $\tau(q_0, o) = \tau'(q_0, o)$, for $q \in Q'_d : \tau(q, o_a) = (\tau'(q, o_a), a)$ and for $q \in Q' : \tau((q, o_1), o_2) = \tau'(q, o_1, o_2)$. The output function is given by: $\lambda(q_0, o) = \lambda'(q_0, o)$, for $q \in Q_d : \lambda(q, o_1) = \lambda'(q, o_1)$ and for $(q, o_1) \in (Q' \times \mathcal{O}_{\mathsf{max}}) : \lambda((q, o_1), o_2) = \lambda'(q, o_1, o_2)$. Mealy machine $M^{\mathfrak{R}(\mathcal{G})}_{a}$ realizes $\Omega_{\mathsf{max}}(\sigma')$.

Special cases. For a Mealy transducer $M = (Q, q_0, u, g)$, a memoryless strategy corresponds to $|Q| = 1$, which the construction trivially preserves. An infinite-memory strategy places no bound on $|Q|$; the construction uses the same Q, so the class is preserved.

□

D.4 Proof of Lemma 3.3

Proof. Because parity objective subsumes all ω-regular objectives, from lemma D.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 D.12, D.13, D.14, and D.15. □