Bellman-Local Lyapunov Barriers for Exact Stationary Nash Learning in Discounted Perfect-Information Stochastic Games

Anonymous Authors[1]

Abstract

We study whether exact stationary Nash equilibria in finite rational two-player discounted perfect-information stochastic games can be reached by universally defined local learning dynamics. We introduce a Bellman-local model in which an update rule may inspect the exact Bellman jet of the current stationary profile—the current policy, exact value vectors, exact one-step continuation values, exact deviation gains, and exact discounted occupancies—and is certified by a Bellman-local Lyapunov witness. Our main result shows that if there exists a universal Bellman-local strict-descent pair whose fixed points are exactly the stationary Nash equilibria on bounded-bit policy grids, then the exact stationary-Nash search problem belongs to PLS; since this problem is already PPAD-complete, this implies the complexity collapse $\mathrm{PPAD} = \mathrm{CLS}$. We further prove a quantitative strengthening: if the same Bellman-local descent admits polynomially bounded witness range and inverse-polynomial one-step progress, then simple iteration computes exact or constant-accuracy stationary Nash in deterministic polynomial time, implying $\mathrm{PPAD} \subseteq \mathrm{FP}$ at sufficiently small constant accuracy. These results identify a sharp frontier for game-theoretic learning in stochastic environments: Bellman locality and exact critic information alone do not suffice for a universal exact descent theory without additional structural assumptions.

1. Introduction

A central goal in multi-agent reinforcement learning and game-theoretic learning is to understand whether strategically interacting agents can reach stable solution concepts through *local* policy updates. In stochastic games, locality naturally means that an update rule should inspect the current policy together with the continuation information generated by that policy: value functions, one-step deviation values, and discounted visitation frequencies. At the same time, exact equilibrium computation in dynamic games is known to be difficult. This raises a basic question: is the failure of universal exact learning in stochastic games primarily a matter of optimization geometry, or is it already a matter of computational complexity?

This paper studies that question for the smallest dynamic class where the general-sum exact equilibrium problem is already hard: finite rational two-player discounted perfect-information stochastic games. These games sit at a particularly interesting frontier. On the one hand, stationary equilibria exist in finite discounted stochastic games by classical results of Shapley (1953); Fink (1964); Takahashi (1964). On the other hand, recent work shows that computing an exact stationary Nash equilibrium in the two-player perfect-information subclass is PPAD-complete (Daskalakis et al., 2023; Hansen & Nie, 2025). By contrast, several positive algorithmic results are known once additional structure is imposed: zero-sum turn-based games admit strong algorithms (Hansen et al., 2013; Sidford et al., 2020; Batziou et al., 2025), while in general-sum Markov games convergence results survive under strong structural assumptions such as potentiality, equilibrium collapse, or near-potentiality (Leonardos et al., 2022; Fox et al., 2022; Kalogiannis & Panageas, 2023; Anagnostides et al., 2024; Maheshwari et al., 2024). The unresolved point is whether there exists a *universal* Bellman-aware local descent principle for exact stationary Nash on the full class.

We answer this question negatively under a standard and natural certification hypothesis. We introduce a model of *Bellman-local* dynamics operating on bounded-bit stationary policy grids. A Bellman-local rule may inspect the *exact Bellman jet* of the current profile—the current policy, exact value vectors, exact one-step continuation values, exact deviation gains, and exact discounted occupancies—and then output the next bounded-bit stationary policy. We pair such an update with a Bellman-local Lyapunov witness and ask whether there can exist a universally defined strict-descent rule whose fixed points are exactly the stationary Nash equi-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

libria. Our main theorem shows that if such a rule existed on the full class of two-player discounted perfect-information stochastic games, then the exact stationary-Nash search problem would belong to PLS. Since that problem is already PPAD-complete, and since CLS = PPAD ∩ PLS (Johnson et al., 1988; Fearnley et al., 2022), this would imply the complexity collapse PPAD = CLS.

The result is genuinely dynamic. Our barrier does *not* reduce to a one-state normal-form embedding, because the update rule is allowed to use exact continuation structure. Indeed, Appendix G shows explicitly that Bellman-local updates can change at a state even when the entire local stage signature at that state is held fixed, provided downstream transition-reward structure changes. Thus the obstruction identified here is not the absence of critic information. Even granting exact Bellman evaluation, a universal strict-descent theory for exact stationary Nash would force an unexpected complexity consequence.

Our contributions are as follows.

- We define *Bellman-local update maps* and *Bellman-local Lyapunov witnesses* on bounded-bit stationary policy grids, and prove that the exact Bellman jet of any bounded-bit stationary profile is itself exactly computable in polynomial time.
- We prove a *main barrier theorem*: if there exists a universal Bellman-local strict-descent pair for exact stationary Nash on the full class $\mathcal{G}_{\mathrm{PI}}$, then the exact stationary-Nash search problem belongs to PLS. Combined with the PPAD-completeness of exact stationary Nash in this class, this yields PPAD = CLS.
- We prove a *quantitative strengthening*: if the same Bellman-local descent rule enjoys polynomially bounded Lyapunov range and inverse-polynomial one-step progress, then simple iteration computes exact or constant-accuracy stationary Nash in deterministic polynomial time. For sufficiently small constant ε, this implies PPAD ⊆ FP.
- We clarify the scope of the result. The theorem is a *frontier theorem*, not a blanket impossibility theorem. It leaves open—and is fully compatible with—positive results on strict structured subclasses, including zero-sum, potential, equilibrium-collapse, and near-potential regimes.

Conceptually, the message is that Bellman locality alone is not the missing ingredient in exact stationary-Nash learning. Exact values, exact continuation-sensitive deviations, and exact discounted occupancies still do not yield a universal local descent principle on the full general-sum class unless one is willing to accept a major complexity collapse. What makes positive results possible is therefore not Bellman locality by itself, but additional structure that turns the induced dynamic search problem into a tractable one.

2. Preliminaries and Bellman-Local Framework

We work with the finite discounted stochastic-game model of Shapley (1953) and its stationary nonzero-sum extension due to Fink (1964); Takahashi (1964). Our complexity-theoretic focus is the class $\mathcal{G}_{\mathrm{PI}}$ of finite rational two-player perfect-information—equivalently, turn-based—discounted stochastic games, for which exact stationary Nash computation forms the sharp PPAD frontier identified by Daskalakis et al. (2023); Hansen & Nie (2025). Full definitions and proofs appear in Sections B to D; here we isolate only the objects needed for the barrier theorem.

2.1. Discounted Perfect-Information Stochastic Games

For a finite set B, let $\Delta(B)$ denote the probability simplex over B. An input game

$$G = (S, S_1, S_2, \{A(s)\}_{s\in S}, P, r_1, r_2, \gamma, \mu) \in \mathcal{G}_{\mathrm{PI}}$$

consists of a finite state space $S = S_1 \dot\cup S_2$, where exactly one player acts at each state, nonempty finite action sets $A(s)$, a transition kernel $P(\cdot \mid s, a) \in \Delta(S)$, bounded stage rewards $r_i(s, a) \in [0, 1]$ for each player $i \in \{1, 2\}$, a discount factor $\gamma \in [0, 1)$, and a reference initial distribution $\mu \in \Delta(S)$. All primitive data are rational. We write $c(s) \in \{1, 2\}$ for the unique controller of state s.

A stationary policy profile is a pair

$$\pi = (\pi_1, \pi_2) \in \Pi(G) := \prod_{s\in S_1} \Delta(A(s)) \times \prod_{s\in S_2} \Delta(A(s)).$$

For $a \in A(s)$, we abbreviate $\pi(a \mid s) := \pi_{c(s)}(a \mid s)$. The induced transition matrix $P^\pi \in \mathbb{R}^{S\times S}$ and reward vectors $r_i^\pi \in \mathbb{R}^S$ are

$$\begin{aligned} P^\pi_{ss'} &:= \sum_{a\in A(s)} \pi(a \mid s)\, P(s' \mid s, a), \\ r_i^\pi(s) &:= \sum_{a\in A(s)} \pi(a \mid s)\, r_i(s, a). \end{aligned} \tag{1}$$

The normalized discounted value vector of player i is

$$V_i^\pi(s) := \mathbb{E}_s^\pi\left[(1-\gamma)\sum_{t=0}^{\infty} \gamma^t r_i(s_t, a_t)\right], \qquad s \in S, \tag{2}$$

and satisfies

$$V_i^\pi = (1-\gamma) r_i^\pi + \gamma P^\pi V_i^\pi. \tag{3}$$

The one-step continuation value of action $a \in A(s)$ is

$$Q_i^\pi(s,a) := (1-\gamma) r_i(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a)\, V_i^\pi(s'), \tag{4}$$

and the corresponding one-step deviation gain is

$$\Delta_i^\pi(s,a) := Q_i^\pi(s,a) - V_i^\pi(s). \tag{5}$$

We also use the discounted occupancy vector

$$d_\mu^\pi := (1-\gamma)\mu^\top (I - \gamma P^\pi)^{-1}. \tag{6}$$

Our equilibrium notion is *statewise* stationary Nash. For a player $i \in \{1,2\}$, define the unilateral best-response value function

$$W_i^\pi(s) := \sup_{\sigma_i} V_i^{(\sigma_i, \pi_{-i})}(s), \qquad s \in S, \tag{7}$$

where the supremum is over all behavioral strategies of player i. For the discounted turn-based class, this supremum is attained by a deterministic stationary best response; see Lemma B.6 in Section B. Following Daskalakis et al. (2023); Hansen & Nie (2025), we say that $\pi \in \Pi(G)$ is a *statewise stationary ε-Nash equilibrium* if

$$W_i^\pi(s) - V_i^\pi(s) \le \varepsilon \qquad \text{for all } i \in \{1,2\},\ s \in S. \tag{8}$$

The case $\varepsilon = 0$ is called a *statewise stationary Nash equilibrium*, and we write

$$\mathrm{SNE}(G) := \left\{ \begin{array}{l} \pi \in \Pi(G) : \\ \pi \text{ is statewise stationary,} \\ \pi \text{ is a Nash equilibrium} \end{array} \right\}.$$

The Bellman objects above characterize exact equilibrium locally.

Proposition 2.1 (Bellman characterization of exact stationary Nash)**.** *For a stationary profile $\pi \in \Pi(G)$, the following are equivalent:*

1. $\pi \in \mathrm{SNE}(G)$;
2. *for every player $i \in \{1,2\}$ and every state $s \in S_i$,*
$$V_i^\pi(s) = \max_{a \in A(s)} Q_i^\pi(s,a);$$
3. *for every player $i \in \{1,2\}$, every $s \in S_i$, and every $a \in A(s)$,*
$$\Delta_i^\pi(s,a) \le 0.$$

Proof. Deferred to Section B; see Proposition B.7. □

2.2. Bounded-Bit Stationary Policy Grids

Our barrier theorem is formulated in the standard Turing bit model. Thus the hypothetical learning dynamics we rule out operate on stationary policies whose coordinates are rational numbers of polynomially bounded encoding length. As throughout, $\langle \cdot \rangle$ denotes binary encoding length.

Fix once and for all a polynomial bit-budget function

$$b : \mathbb{N} \to \mathbb{N}, \qquad b_G := b(\langle G \rangle).$$

Let $\Pi_b(G)$ denote the set of stationary profiles whose every action probability has reduced numerator and denominator strictly smaller than 2^{b_G}. Formally,

$$\Pi_b(G) := \Pi_{b_G}(G),$$

where $\Pi_{b_G}(G)$ is the bounded-bit grid constructed in Section D. We also write

$$\mathrm{SNE}_b(G) := \mathrm{SNE}(G) \cap \Pi_b(G)$$

for the exact stationary equilibria representable on this grid.

The set $\Pi_b(G)$ admits a canonical encoding as a finite subset of a Boolean hypercube. Specifically, Section D constructs a valid encoding set

$$\mathcal{V}_b(G) \subseteq \{0,1\}^{\ell_b(G)}, \qquad \ell_b(G) = 2b_G \sum_{s \in S} |A(s)|,$$

together with a bijection

$$\mathrm{enc}_{G,b} : \Pi_b(G) \to \mathcal{V}_b(G)$$

and inverse decoding map

$$\mathrm{dec}_{G,b} : \mathcal{V}_b(G) \to \Pi_b(G).$$

The ambient cube $\{0,1\}^{\ell_b(G)}$ also contains invalid strings; these are handled later by a canonical dummy sink.

Lemma 2.2 (Finite encoded policy grid)**.** *For every game $G \in \mathcal{G}_{\mathrm{PI}}$, the set $\Pi_b(G)$ is finite, canonically encoded by $\mathcal{V}_b(G)$, and satisfies*

$$|\Pi_b(G)| = |\mathcal{V}_b(G)| \le 2^{\ell_b(G)}.$$

Moreover, validity testing, encoding, and decoding are all executable in time polynomial in $\langle G \rangle + b_G$.

Proof. Deferred to Section D; see Lemma D.8 and Proposition D.9. □

Remark 2.3. The bounded-bit grid is part of the computational model, not a numerical approximation device. Our impossibility result is conditional on the existence of a *locally tractable* Bellman-aware update rule whose iterates and certificates are representable with polynomially many bits.

2.3. Bellman-Local Dynamics and Lyapunov Witnesses

The local information available to the hypothetical learning dynamics is the exact Bellman geometry of the current stationary profile.

Definition 2.4 (Bellman jet)**.** For $G \in \mathcal{G}_{\mathrm{PI}}$ and $\pi \in \Pi(G)$, the *Bellman jet* of π is

$$J_G(\pi) := \Big(\pi,\ V_1^\pi, V_2^\pi,\ Q_1^\pi, Q_2^\pi,\ \Delta_1^\pi, \Delta_2^\pi,\ d_\mu^\pi\Big).$$

Thus Bellman-locality grants access not only to the current policy, but also to exact policy evaluation, exact one-step continuation values, exact unilateral deviation gains, and exact discounted visitation weights.

Definition 2.5 (Bellman-local update map)**.** A family $F = \{F_G\}_{G \in \mathcal{G}_{\mathrm{PI}}}$ is a *Bellman-local update map* (relative to the bit budget b) if, for every $G \in \mathcal{G}_{\mathrm{PI}}$,

$$F_G : \Pi_b(G) \to \Pi_b(G),$$

and there exists a deterministic algorithm running in time polynomial in $\langle G \rangle + b_G$ that, on input a rational encoding of the Bellman jet $J_G(\pi)$, outputs $\mathrm{enc}_{G,b}(F_G(\pi))$.

Definition 2.6 (Bellman-local Lyapunov witness)**.** A family $L = \{L_G\}_{G \in \mathcal{G}_{\mathrm{PI}}}$ is a *Bellman-local Lyapunov witness* (relative to b) if, for every game $G \in \mathcal{G}_{\mathrm{PI}}$,

$$L_G : \Pi_b(G) \to \mathbb{Q},$$

and $L_G(\pi)$ is computable in time polynomial in $\langle G \rangle + b_G$ from a rational encoding of $J_G(\pi)$.

Definition 2.7 (Strict Bellman-Lyapunov descent pair)**.** A pair (F, L) of Bellman-local families is a *strict Bellman-Lyapunov descent pair* if, for every $G \in \mathcal{G}_{\mathrm{PI}}$,

1. $F_G(\pi) = \pi$ for every $\pi \in \mathrm{SNE}_b(G)$;
2. $L_G(F_G(\pi)) < L_G(\pi)$ for every $\pi \in \Pi_b(G) \setminus \mathrm{SNE}_b(G)$.

The next lemma is what makes the Bellman-local model strong enough for the barrier theorem: the update rule is allowed to use the *exact* Bellman jet of the current profile.

Lemma 2.8 (Exact Bellman-jet computability)**.** *For every game $G \in \mathcal{G}_{\mathrm{PI}}$ and every $\pi \in \Pi_b(G)$, the Bellman jet $J_G(\pi)$ is exactly computable in time polynomial in $\langle G \rangle + b_G$. Moreover, every coordinate of $J_G(\pi)$ has encoding length polynomial in $\langle G \rangle + b_G$.*

Proof. Deferred to Section C; see Theorem C.6 and Corollary C.7. □

Finally, because $\Pi_b(G)$ is finite, strict Bellman-Lyapunov descent cannot cycle outside the target equilibrium set.

Proposition 2.9 (Finite-grid descent principle)**.** *Let (F, L) be a strict Bellman-Lyapunov descent pair. Then for every $G \in \mathcal{G}_{\mathrm{PI}}$, the forward orbit of F_G from any initial point in $\Pi_b(G)$ reaches $\mathrm{SNE}_b(G)$ in finitely many steps and remains there forever.*

Proof. Deferred to Section D; see Proposition D.13 and Corollary D.14. □

3. Main Barrier Theorem

We now state the central barrier theorem. The search problem of interest is

$$\mathsf{StatNE}_{\mathrm{PI}} : \begin{cases} \text{input } G \in \mathcal{G}_{\mathrm{PI}}, \\ \text{output any } \pi \in \mathrm{SNE}(G). \end{cases}$$

The point of the theorem is not merely that exact stationary Nash equilibrium is computationally hard. Rather, it shows that if one postulates a *universally convergent*, *locally tractable*, and *Bellman-aware* descent dynamic for exact stationary Nash, then the resulting search problem falls inside the local-search class PLS in the standard sense of Johnson et al. (1988). Since exact stationary Nash in two-player discounted perfect-information stochastic games is already PPAD-complete (Daskalakis et al., 2023; Hansen & Nie, 2025), this yields a complexity-collapse consequence.

Theorem 3.1 (Bellman-local strict descent implies PLS membership)**.** *Fix a polynomial bit-budget function b, and suppose there exists a strict Bellman-Lyapunov descent pair (F, L) relative to b on $\mathcal{G}_{\mathrm{PI}}$ in the sense of Definition 2.7. Then*

$$\mathsf{StatNE}_{\mathrm{PI}} \in \mathrm{PLS}.$$

More precisely, the bounded-bit exact equilibrium problem

$$\mathsf{GridSNE}^b_{\mathrm{PI}} : \begin{cases} \textit{input } G \in \mathcal{G}_{\mathrm{PI}}, \\ \textit{output any } \pi \in \mathrm{SNE}_b(G) \end{cases}$$

belongs to PLS*, and therefore so does* $\mathsf{StatNE}_{\mathrm{PI}}$.

Proof sketch. Fix an input game G. We construct a local-search instance on the full Boolean hypercube

$$\mathcal{X}_b(G) = \{0,1\}^{\ell_b(G)}$$

that canonically encodes the bounded-bit stationary policy grid $\Pi_b(G)$; see Section 2.2. Valid strings $x \in \mathcal{V}_b(G) \subseteq \mathcal{X}_b(G)$ decode to unique policies $\pi = \mathrm{dec}_{G,b}(x) \in \Pi_b(G)$, while invalid strings are redirected to a canonical dummy encoding $x^\circ_{G,b}$.

The designated successor of a valid encoding is obtained by applying the Bellman-local update map:

$$x = \mathrm{enc}_{G,b}(\pi) \quad \longmapsto \quad S_G(x) := \mathrm{enc}_{G,b}(F_G(\pi)).$$

The objective is an integer-valued reparametrization of the Lyapunov witness. Since $L_G(\pi)$ is rational-valued, one first proves that Bellman-locality implies a polynomial bound on the encoding length of every witness value. One then applies an order-preserving integerization to the reversed witness values, thereby obtaining an integer objective Φ_G that strictly increases whenever L_G strictly decreases. Invalid strings are assigned objective value 0, while the dummy encoding receives a strictly larger objective, so no invalid string can be locally optimal.

The construction is polynomial-time because exact Bellman jets are polynomial-time computable on the bounded-bit grid (Lemma 2.8). Hence the successor circuit may use the *exact* value vectors, *exact* one-step continuation values, *exact* deviation gains, and *exact* discounted occupancies of the current profile. This is the sense in which the barrier is genuinely Bellman-local rather than merely stage-local.

It remains to characterize local optima. By construction, invalid strings are never local optima. If $x \in \mathcal{V}_b(G)$ decodes to a policy $\pi \notin \mathrm{SNE}_b(G)$, then strict Bellman-Lyapunov descent implies

$$L_G(F_G(\pi)) < L_G(\pi),$$

and therefore the integerized objective strictly improves along the designated successor:

$$\Phi_G(S_G(x)) > \Phi_G(x).$$

Thus every local optimum of (S_G, Φ_G) must decode to an element of $\mathrm{SNE}_b(G)$. This yields $\mathsf{GridSNE}^b_{\mathrm{PI}} \in \mathrm{PLS}$. Since $\mathrm{SNE}_b(G) \subseteq \mathrm{SNE}(G)$, the same local-search construction solves $\mathsf{StatNE}_{\mathrm{PI}}$. The full proof is given in Section E. □

Corollary 3.2 (Complexity collapse)**.** *Under the hypothesis of Theorem 3.1,*

$$\mathrm{PPAD} = \mathrm{CLS}.$$

Proof. By Theorem 3.1, $\mathsf{StatNE}_{\mathrm{PI}} \in \mathrm{PLS}$. On the other hand, computing an exact stationary Nash equilibrium in two-player discounted perfect-information stochastic games is PPAD-complete (Daskalakis et al., 2023; Hansen & Nie, 2025). Hence

$$\mathrm{PPAD} \subseteq \mathrm{PLS}.$$

Using the class identity

$$\mathrm{CLS} = \mathrm{PPAD} \cap \mathrm{PLS}$$

proved by Fearnley et al. (2022), we obtain

$$\mathrm{PPAD} = \mathrm{PPAD} \cap \mathrm{PLS} = \mathrm{CLS}.$$ □

Remark 3.3 (Why the theorem is genuinely dynamic). Theorem 3.1 is strictly stronger than a one-state embedding of a normal-form barrier. The successor map is allowed to depend on the full Bellman jet $J_G(\pi)$, including continuation values and discounted occupancy weights, which vary with downstream transition structure even when stage rewards at the current state are held fixed. Thus the obstruction is not that the dynamic lacks access to critic information; the obstruction is computational.

4. Quantitative Strengthening

Theorem 3.1 rules out universal Bellman-local *strict* descent unless PPAD = CLS. We now show that a stronger quantitative hypothesis—polynomially bounded witness range together with inverse-polynomial one-step progress—would force deterministic polynomial-time computation. For $\varepsilon \geq 0$, let $\mathrm{SNE}^{\varepsilon}_b(G)$ denote the set of statewise stationary ε-Nash equilibria in $\Pi_b(G)$. Fix a polynomial bit-budget function b, and abbreviate

$$b_G := b(\langle G \rangle), \qquad n_G := \langle G \rangle + b_G.$$

Suppose there exist Bellman-local families (F, L), polynomials B, q, and a constant $\varepsilon \geq 0$ such that for every $G \in \mathcal{G}_{\mathrm{PI}}$:

$$0 \leq L_G(\pi) \leq B(n_G) \quad \forall \pi \in \Pi_b(G), \tag{9}$$

$$F_G(\pi) = \pi \qquad \forall \pi \in \mathrm{SNE}^{\varepsilon}_b(G), \tag{10}$$

$$L_G(\pi) - L_G(F_G(\pi)) \geq \frac{1}{q(n_G)} \qquad \forall \pi \in \Pi_b(G) \setminus \mathrm{SNE}^{\varepsilon}_b(G). \tag{11}$$

The next theorem says that these assumptions already imply a polynomial-time algorithm.

Theorem 4.1 (Quantitative Bellman descent implies deterministic polynomial time)**.** *Assume (9)–(11) hold for some fixed constant $\varepsilon \geq 0$. Then*

$$\mathsf{GridSNE}^{b,\varepsilon}_{\mathrm{PI}} \in \mathrm{FP}.$$

In particular, when $\varepsilon = 0$,

$$\mathsf{StatNE}_{\mathrm{PI}} \in \mathrm{FP}.$$

Proof sketch. Start from the canonical dummy policy $\pi^{\circ} \in \Pi_b(G)$, and iterate

$$\pi^{(t+1)} := F_G(\pi^{(t)}).$$

Define the time horizon

$$T_G := B(n_G)\, q(n_G) + 1.$$

Because F is Bellman-local and exact Bellman jets are polynomial-time computable on $\Pi_b(G)$ (Lemma 2.8), each iterate $\pi^{(t+1)}$ is computable exactly in time polynomial in $\langle G \rangle$. Since T_G is also polynomial in $\langle G \rangle$, the orbit can be simulated for T_G steps in deterministic polynomial time.

If none of the first T_G iterates belonged to $\mathrm{SNE}_b^{\varepsilon}(G)$, then summing the inverse-polynomial progress bound (11) along the orbit would yield

$$L_G(\pi^{(0)}) - L_G(\pi^{(T_G)}) \geq \frac{T_G}{q(n_G)} > B(n_G),$$

contradicting the range bound (9). Hence $\pi^{(T_G)} \in \mathrm{SNE}_b^{\varepsilon}(G)$. This proves $\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon} \in \mathrm{FP}$. When $\varepsilon = 0$, the output lies in $\mathrm{SNE}_b(G) \subseteq \mathrm{SNE}(G)$, so the same algorithm solves $\mathsf{StatNE}_{\mathrm{PI}}$. The full proof appears in Section F. □

The exact case already implies that universal inverse-polynomial Bellman descent would collapse the entire PPAD frontier to deterministic polynomial time, because exact stationary Nash in this class is PPAD-complete (Daskalakis et al., 2023; Hansen & Nie, 2025). The approximate case admits an explicit constant-threshold corollary.

Corollary 4.2 (Explicit constant-accuracy collapse)**.** *Let*

$$\bar{\varepsilon} := \frac{3 - 2\sqrt{2}}{288}.$$

Fix any constant $0 \leq \varepsilon < \bar{\varepsilon}$*. If* (9)–(11) *hold with target set* $\mathrm{SNE}_b^{\varepsilon}(G)$*, then*

$$\mathrm{PPAD} \subseteq \mathrm{FP}.$$

Proof. By Theorem 4.1, $\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon} \in \mathrm{FP}$, and hence so is the unrestricted ε-stationary-Nash problem, since every output in $\mathrm{SNE}_b^{\varepsilon}(G)$ is also an unrestricted statewise stationary ε-Nash equilibrium. Hansen and Nie prove that for every

$$0 \leq \varepsilon < \frac{3 - 2\sqrt{2}}{288},$$

computing such an ε-equilibrium in two-player $1/2$-discounted perfect-information stochastic games is PPAD-hard (Hansen & Nie, 2025, Theorem 2). Therefore the assumed quantitative Bellman descent would imply deterministic polynomial-time algorithms for all problems in PPAD. □

Remark 4.3. Section 3 rules out universal Bellman-local descent unless PPAD = CLS. The present section rules out a substantially stronger possibility: a Bellman-local descent dynamic with *uniform* inverse-polynomial progress and polynomial witness range would not merely yield a local-search interpretation—it would directly compute a PPAD-hard equilibrium in deterministic polynomial time.

5. Scope, Frontier, and Why Structure Matters

Theorems 3.1 and 4.1 are *universal* statements: they assume Bellman-local descent principles on the full class $\mathcal{G}_{\mathrm{PI}}$ and derive consequences for the unrestricted exact or constant-accuracy stationary-Nash search problems. They therefore do *not* say that Bellman-local learning is impossible on every structured family of stochastic games. Rather, they isolate a frontier: without additional structure, one cannot hope for a single exact Bellman-Lyapunov principle that solves stationary Nash across all two-player discounted perfect-information games. Appendix H makes this restriction principle formal.

This distinction is exactly what separates the negative result of this paper from the existing positive literature. In zero-sum turn-based stochastic games, strong algorithms exist for values and ε-optimal strategies, and the problem exhibits contraction and monotonicity structure unavailable in the general-sum exact stationary-Nash setting (Hansen et al., 2013; Sidford et al., 2020; Batziou et al., 2025). In equilibrium-collapse regimes, approximate Nash learning becomes tractable under additional collapse or single-controller assumptions (Kalogiannis & Panageas, 2023; Anagnostides et al., 2024). In Markov potential and near-potential games, one obtains global or last-iterate convergence, or approximate Lyapunov control, from special geometric structure that is absent in the full class $\mathcal{G}_{\mathrm{PI}}$ (Leonardos et al., 2022; Fox et al., 2022; Maheshwari et al., 2024). What our theorem rules out is not Bellman-local learning per se, but a *universal* exact Bellman-local strict-descent theory for stationary Nash without such structure.

Conceptually, the frontier suggested by our results is the following. Positive theory survives when the game class carries additional geometry—contraction, potentiality, equilibrium collapse, or a near-potential surrogate—that turns Bellman-local updates into a tractable search process. The unrestricted exact stationary-Nash problem appears to sit beyond that frontier: Bellman locality and exact critic information alone do not suffice to produce a universal local descent principle without implying a complexity collapse.

Remark 5.1 (Open problem: removing the witness). Theorem 3.1 assumes an explicit Bellman-local Lyapunov witness. An important next question is whether this certificate is actually necessary. Suppose a Bellman-local family $F = \{F_G\}_{G \in \mathcal{G}_{\mathrm{PI}}}$ has the property that, for every $G \in \mathcal{G}_{\mathrm{PI}}$, every orbit of F_G on $\Pi_b(G)$ eventually reaches $\mathrm{SNE}_b(G)$. Must this already imply that $\mathsf{StatNE}_{\mathrm{PI}} \in \mathrm{PLS}$, or at least that exact stationary Nash admits a comparable local-search barrier? A positive answer would show that the obstruction identified here is intrinsic to Bellman-local convergence itself, not merely to the existence of an explicit witness.

References

Anagnostides, I., Panageas, I., Farina, G., and Sandholm, T. Optimistic policy gradient in multi-player markov games with a single controller: Convergence beyond the Minty property. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 9451–9459, 2024. doi: 10.1609/aaai.v38i9.28799. URL https://ojs.aaai.org/index.php/AAAI/article/view/28799.

Bareiss, E. H. Sylvester's identity and multistep integer-preserving gaussian elimination. *Mathematics of Computation*, 22(103):565–578, 1968. doi: 10.1090/S0025-5718-1968-0226829-0. URL https://doi.org/10.1090/S0025-5718-1968-0226829-0.

Batziou, E., Fearnley, J., Gordon, S., Mehta, R., and Savani, R. Monotone contractions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, pp. 507–517. ACM, 2025. doi: 10.1145/3717823.3718175. URL https://doi.org/10.1145/3717823.3718175.

Daskalakis, C., Golowich, N., and Zhang, K. The complexity of markov equilibrium in stochastic games. In Neu, G. and Rosasco, L. (eds.), *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 4180–4234. PMLR, 2023. URL https://proceedings.mlr.press/v195/daskalakis23a.html.

Fearnley, J., Goldberg, P. W., Hollender, A., and Savani, R. The complexity of gradient descent: CLS = PPAD ∩ PLS. *Journal of the ACM*, 70(1):7:1–7:74, 2022. doi: 10.1145/3568163. URL https://doi.org/10.1145/3568163.

Fink, A. M. Equilibrium in a stochastic n-person game. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)*, 28(1):89–93, 1964. doi: 10.32917/hmj/1206139508. URL https://doi.org/10.32917/hmj/1206139508.

Fox, R., McAleer, S. M., Overman, W., and Panageas, I. Independent natural policy gradient always converges in markov potential games. In Camps-Valls, G., Ruiz, F. J. R., and Valera, I. (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 4414–4425. PMLR, 2022. URL https://proceedings.mlr.press/v151/fox22a.html.

Hansen, K. A. and Nie, X. On the complexity of stationary Nash equilibria in discounted perfect information stochastic games. *arXiv preprint arXiv:2510.11550*, 2025. doi: 10.48550/arXiv.2510.11550. URL https://arxiv.org/abs/2510.11550.

Hansen, T. D., Miltersen, P. B., and Zwick, U. Strategy iteration is strongly polynomial for 2-player turn-based stochastic games with a constant discount factor. *Journal of the ACM*, 60(1):1:1–1:16, 2013. doi: 10.1145/2432622.2432623. URL https://doi.org/10.1145/2432622.2432623.

Johnson, D. S., Papadimitriou, C. H., and Yannakakis, M. How easy is local search? *Journal of Computer and System Sciences*, 37(1):79–100, 1988. doi: 10.1016/0022-0000(88)90046-3. URL https://doi.org/10.1016/0022-0000(88)90046-3.

Kalogiannis, F. and Panageas, I. Zero-sum polymatrix markov games: Equilibrium collapse and efficient computation of Nash equilibria. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/bcdcd565f83a8a6681a8269d325a5304-Abstract-Confere
html.

Kannan, R. and Bachem, A. Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix. *SIAM Journal on Computing*, 8(4):499–507, 1979. doi: 10.1137/0208040. URL https://doi.org/10.1137/0208040.

Leonardos, S., Overman, W., Panageas, I., and Piliouras, G. Global convergence of multi-agent policy gradient in markov potential games. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=gfwON7rAm4.

Maheshwari, C., Wu, M., and Sastry, S. S. Convergence of decentralized actor-critic algorithm in general-sum markov games. *IEEE Control Systems Letters*, 8:2643–2648, 2024. doi: 10.1109/LCSYS.2024.3510193. URL https://doi.org/10.1109/LCSYS.2024.3510193.

Shapley, L. S. Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095–1100, 1953. doi: 10.1073/pnas.39.10.1095. URL https://doi.org/10.1073/pnas.39.10.1095.

Sidford, A., Wang, M., Yang, L. F., and Ye, Y. Solving discounted stochastic two-player games with near-optimal time and sample complexity. In Chiappa, S. and Calandra, R. (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2992–3002. PMLR,

2020. URL https://proceedings.mlr.press/v108/sidford20a.html.

Takahashi, M. Equilibrium points of stochastic noncooperative n-person games. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)*, 28(1):95–99, 1964. doi: 10.32917/hmj/1206139509. URL https://doi.org/10.32917/hmj/1206139509.

A. Complexity Preliminaries

This appendix fixes the search-problem formalism and the exact complexity facts used in the proof of the main barrier theorem. We only require a circuit-style presentation of PLS, together with two external facts:

1. exact stationary Nash equilibrium in two-player discounted perfect-information stochastic games is PPAD-complete (Daskalakis et al., 2023; Hansen & Nie, 2025);
2. $\mathrm{CLS} = \mathrm{PPAD} \cap \mathrm{PLS}$ (Fearnley et al., 2022).

No direct reduction to PPAD or CLS is constructed in this paper; the only class we reduce to explicitly is PLS.

A.1. Search Problems

Definition A.1 (Polynomially balanced search problem)**.** A *search problem* is specified by a pair $(\mathcal{I}, R)$, where $\mathcal{I} \subseteq \{0,1\}^*$ is a set of instances and

$$R \subseteq \{(x, y) \in \{0,1\}^* \times \{0,1\}^* : x \in \mathcal{I}\}$$

is a binary relation of valid solutions, together with a polynomial p such that

$$(x, y) \in R \implies |y| \le p(|x|).$$

For $x \in \mathcal{I}$, we write

$$\mathrm{Sol}_R(x) := \{y \in \{0,1\}^* : (x, y) \in R\}.$$

The search problem is *total* if $\mathrm{Sol}_R(x) \neq \varnothing$ for every $x \in \mathcal{I}$.

Remark A.2. We follow the standard TFNP convention that solutions are represented by binary strings of polynomially bounded length. In our applications, a solution string encodes either a rational stationary policy or a bounded-bit valid policy-grid encoding.

A.2. A Circuit Presentation of PLS

We use the following successor-and-potential presentation of polynomial local search. It is equivalent, up to standard polynomial-time intertranslation, to the original neighborhood-and-cost formulation of Johnson et al. (1988). We use this form because the proof of Theorem E.10 constructs exactly a polynomial-time successor map and a polynomially bounded integer objective.

Definition A.3 (Circuit local-search instance)**.** A *circuit local-search instance* is a tuple

$$\mathcal{L} = (m, x^{\mathrm{init}}, S, V),$$

where:

1. $m \in \mathbb{N}$ is the search dimension;
2. $x^{\mathrm{init}} \in \{0,1\}^m$ is an explicitly given initial point;
3. $S : \{0,1\}^m \to \{0,1\}^m$ is a polynomial-time computable successor map;
4. $V : \{0,1\}^m \to \mathbb{Z}_{\ge 0}$ is a polynomial-time computable objective function whose output bit length is polynomially bounded in the instance size.

A point $x \in \{0,1\}^m$ is a *local optimum* of $\mathcal{L}$ if

$$V(x) \ge V(S(x)).$$

Remark A.4. The initial point x^{init} is included to align with the classical presentation of PLS (Johnson et al., 1988). It plays no role in the correctness of our reduction, because local optima are defined directly from the successor map and the objective on a finite domain.

Lemma A.5 (Existence of local optima)**.** *Every circuit local-search instance has at least one local optimum.*

Proof. Let $\mathcal{L} = (m, x^{\text{init}}, S, V)$ be a circuit local-search instance. Since $\{0,1\}^m$ is finite and V is integer-valued, there exists $x^\star \in \{0,1\}^m$ maximizing V globally. Then

$$V(S(x^\star)) \le V(x^\star),$$

so $x^\star$ is a local optimum. □

Definition A.6 (PLS)**.** A total search problem $(\mathcal{I}, R)$ belongs to PLS if there exist polynomial-time computable maps

$$K : \mathcal{I} \to \{\text{circuit local-search instances}\}, \qquad D : \{(x,z) : x \in \mathcal{I}\} \to \{0,1\}^*,$$

such that for every input $x \in \mathcal{I}$:

1. $K(x) = \mathcal{L}_x$ is a circuit local-search instance;
2. for every local optimum z of $\mathcal{L}_x$,
$$D(x,z) \in \mathrm{Sol}_R(x).$$

Remark A.7. Definition A.6 is the only PLS formalism needed in this paper. It is equivalent to the standard definition based on a polynomial-time initial feasible point, a neighborhood function, and an objective function (Johnson et al., 1988).

A.3. The Stationary-Nash Search Problems Used in This Paper

We now state the two search problems that appear in Sections 3 and E.

Definition A.8 (Exact stationary Nash search problem)**.** The search problem

$$\mathsf{StatNE}_{\mathrm{PI}}$$

has as input a game $G \in \mathcal{G}_{\mathrm{PI}}$, encoded in the standard rational Turing representation from Section B. A solution is any binary string encoding a rational stationary policy profile $\pi \in \Pi(G)$ such that

$$\pi \in \mathrm{SNE}(G).$$

Definition A.9 (Bounded-bit exact stationary Nash search problem)**.** Fix a polynomial bit-budget function b. The search problem

$$\mathsf{GridSNE}^b_{\mathrm{PI}}$$

has as input a game $G \in \mathcal{G}_{\mathrm{PI}}$. A solution is any valid encoding $x \in \mathcal{V}_b(G)$ such that

$$\mathrm{dec}_{G,b}(x) \in \mathrm{SNE}_b(G).$$

Equivalently, one may regard the output as the decoded policy profile $\pi \in \mathrm{SNE}_b(G)$.

Remark A.10. The bounded-bit problem $\mathsf{GridSNE}^b_{\mathrm{PI}}$ need not be total a priori, since $\mathrm{SNE}_b(G)$ could be empty for a given game G. In our main theorem, totality is supplied conditionally by the existence of a strict Bellman-Lyapunov descent pair, via Proposition D.13.

A.4. External Complexity Facts

The proof of the barrier theorem uses only the following two facts.

Theorem A.11 (Exact stationary Nash is PPAD-complete)**.** *The search problem* $\mathsf{StatNE}_{\mathrm{PI}}$ *is* PPAD*-complete.*

Proof. Daskalakis, Golowich, and Zhang prove PPAD-hardness for stationary Markov equilibrium computation in stochastic games, including the two-player turn-based discounted setting (Daskalakis et al., 2023). Hansen and Nie prove that computing an exact stationary Nash equilibrium in two-player discounted perfect-information stochastic games lies in PPAD; in particular, their result implies that the problem is polynomially balanced and total in the standard Turing representation. Combining the two results yields PPAD-completeness (Hansen & Nie, 2025). □

Theorem A.12 ($\mathrm{CLS} = \mathrm{PPAD} \cap \mathrm{PLS}$)**.**

$$\mathrm{CLS} = \mathrm{PPAD} \cap \mathrm{PLS}.$$

Proof. This is Theorem 1.1 of Fearnley et al. (2022). □

Remark A.13. We do not require a direct formal definition of CLS in this paper. The only property of CLS used in the main argument is the class identity in Theorem A.12.

B. Discounted Perfect-Information Stochastic Games

This appendix fixes the stochastic-game model used throughout the paper. We work with the finite discounted model introduced by Shapley (1953) and its nonzero-sum stationary-equilibrium extension due to Fink (1964); Takahashi (1964). Our complexity focus is the two-player perfect-information (equivalently, turn-based) subclass, which is the class for which stationary Nash computation is now known to be PPAD-complete (Daskalakis et al., 2023; Hansen & Nie, 2025).

For a finite set B, let

$$\Delta(B) \;:=\; \Big\{ x \in \mathbb{R}^B_{\geq 0} : \sum_{b \in B} x(b) = 1 \Big\}.$$

All vectors are column vectors unless explicitly transposed, and $\mathbf{1}$ denotes the all-ones vector of the appropriate dimension.

B.1. Model

Definition B.1 (Discounted perfect-information stochastic game)**.** A *discounted perfect-information stochastic game* is a tuple

$$G = \Big(S, S_1, S_2, \{A(s)\}_{s \in S}, P, r_1, r_2, \gamma, \mu \Big)$$

with the following components.

1. S is a finite state space, partitioned as
$$S = S_1 \dot{\cup} S_2,$$
where S_i is the set of states controlled by player $i \in \{1, 2\}$. We write $c(s) \in \{1, 2\}$ for the unique controller of state s, i.e., $c(s) = i$ iff $s \in S_i$.
2. For each state $s \in S$, $A(s)$ is a nonempty finite action set. Since the game is turn-based, exactly one player acts at each state.
3. $P(\cdot \mid s, a) \in \Delta(S)$ is the transition law from state s when action $a \in A(s)$ is played.
4. For each player $i \in \{1, 2\}$,
$$r_i : S \times \bigcup_{s \in S} A(s) \to [0, 1]$$
is the stage-reward function, with $r_i(s, a)$ defined whenever $a \in A(s)$.
5. $\gamma \in [0, 1)$ is the discount factor.
6. $\mu \in \Delta(S)$ is a reference initial-state distribution.

Throughout the paper, all primitive data of G are assumed to be rational.

Remark B.2. The distribution μ is included because our Bellman-local objects will use discounted occupancy measures. The equilibrium notion itself is *statewise* and does not depend on μ.

A finite history of length t is a sequence

$$h_t = (s_0, a_0, s_1, a_1, \ldots, s_{t-1}, a_{t-1}, s_t)$$

such that $a_\tau \in A(s_\tau)$ for every $\tau < t$, and $P(s_{\tau+1} \mid s_\tau, a_\tau) > 0$ whenever the history has positive probability. We write $\mathrm{last}(h_t) = s_t$.

A *behavioral strategy* for player i is a map σ_i which, for every finite history h_t with $\mathrm{last}(h_t) \in S_i$, assigns a distribution

$$\sigma_i(h_t) \in \Delta(A(\mathrm{last}(h_t))).$$

A *stationary policy* for player i is a map

$$\pi_i : S_i \to \bigcup_{s \in S_i} \Delta(A(s)), \qquad \pi_i(s) \in \Delta(A(s)).$$

We write

$$\Pi_i(G) := \prod_{s \in S_i} \Delta(A(s)), \qquad \Pi(G) := \Pi_1(G) \times \Pi_2(G)$$

for the stationary-policy sets of player i and of the full game, respectively. For $\pi = (\pi_1, \pi_2) \in \Pi(G)$, we use the shorthand

$$\pi(a \mid s) := \pi_{c(s)}(a \mid s), \qquad a \in A(s).$$

Any behavioral profile $\sigma = (\sigma_1, \sigma_2)$ and initial state $s \in S$ induce a unique probability measure on infinite plays; we denote the corresponding expectation by $\mathbb{E}_s^\sigma[\cdot]$. We use the normalized discounted payoff convention

$$V_i^\sigma(s) := \mathbb{E}_s^\sigma\left[(1-\gamma)\sum_{t=0}^{\infty} \gamma^t\, r_i(s_t, a_t)\right], \qquad i \in \{1,2\}. \tag{12}$$

Since $r_i \in [0,1]$, the normalization implies $V_i^\sigma(s) \in [0,1]$ for all i, s, σ. For an initial distribution $\mu \in \Delta(S)$, define

$$V_i^\sigma(\mu) := \sum_{s \in S} \mu(s) V_i^\sigma(s).$$

B.2. Statewise Stationary Equilibrium Notions

Definition B.3 (Statewise stationary ε-Nash equilibrium)**.** Let $\varepsilon \geq 0$. A stationary profile $\pi \in \Pi(G)$ is a *statewise stationary ε-Nash equilibrium* if, for every player $i \in \{1,2\}$ and every state $s \in S$,

$$V_i^\pi(s) \;\geq\; \sup_{\sigma_i} V_i^{(\sigma_i, \pi_{-i})}(s) - \varepsilon, \tag{13}$$

where the supremum is over all behavioral strategies σ_i of player i. When $\varepsilon = 0$, we call π a *statewise stationary Nash equilibrium*.

Remark B.4. Definition B.3 is the statewise analogue of the standard discounted ε-Nash condition. It is the appropriate notion for our Bellman-local barrier results because the latter are local in state space; it also matches the perfect-information hardness frontier studied in Daskalakis et al. (2023); Hansen & Nie (2025).

For a stationary profile π, define the *stationary best-response value* of player i against π_{-i} by

$$W_i^\pi(s) := \sup_{\sigma_i} V_i^{(\sigma_i, \pi_{-i})}(s), \qquad s \in S. \tag{14}$$

Then π is a statewise stationary ε-Nash equilibrium iff

$$W_i^\pi(s) - V_i^\pi(s) \leq \varepsilon \qquad \text{for all } i \in \{1,2\},\ s \in S.$$

B.3. Bellman Objects Associated with a Stationary Profile

Fix a stationary profile $\pi \in \Pi(G)$. Define the induced transition matrix $P^\pi \in \mathbb{R}^{S \times S}$ and reward vectors $r_i^\pi \in \mathbb{R}^S$ by

$$P_{ss'}^\pi := \sum_{a \in A(s)} \pi(a \mid s)\, P(s' \mid s, a), \tag{15}$$

$$r_i^\pi(s) := \sum_{a \in A(s)} \pi(a \mid s)\, r_i(s, a). \tag{16}$$

The corresponding discounted value vector $V_i^\pi \in \mathbb{R}^S$ is defined by

$$V_i^\pi(s) = \mathbb{E}_s^\pi\left[(1-\gamma)\sum_{t=0}^{\infty}\gamma^t r_i(s_t,a_t)\right]. \tag{17}$$

For each player i, state $s \in S$, and action $a \in A(s)$, define the one-step continuation quantity

$$Q_i^\pi(s,a) := (1-\gamma)\, r_i(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a)\, V_i^\pi(s'). \tag{18}$$

The associated deviation gain is

$$\Delta_i^\pi(s,a) := Q_i^\pi(s,a) - V_i^\pi(s). \tag{19}$$

When $s \in S_i$, $\Delta_i^\pi(s,a)$ is the value gain to player i from deviating only at the current visit to s, taking action a, and thereafter reverting to π. When $s \in S_{-i}$, the quantity $\Delta_i^\pi(s,a)$ measures player i's continuation sensitivity to the opponent's action at s; we keep these terms because our Bellman-local objects will be allowed to depend on the full local continuation geometry.

Finally, the discounted state-occupancy vector of π under μ is

$$d_\mu^\pi := (1-\gamma)\sum_{t=0}^{\infty}\gamma^t \mu^\top (P^\pi)^t \ \in \ \mathbb{R}^{1\times S}. \tag{20}$$

Proposition B.5 (Basic Bellman identities)**.** *For every stationary profile $\pi \in \Pi(G)$ and each player $i \in \{1,2\}$, the following hold.*

1. *The matrix $I - \gamma P^\pi$ is invertible, and*

$$V_i^\pi = (1-\gamma) r_i^\pi + \gamma P^\pi V_i^\pi = (1-\gamma)(I-\gamma P^\pi)^{-1} r_i^\pi. \tag{21}$$

2. *For every state $s \in S$,*

$$V_i^\pi(s) = \sum_{a \in A(s)} \pi(a \mid s)\, Q_i^\pi(s,a). \tag{22}$$

3. *The occupancy vector satisfies*

$$d_\mu^\pi = (1-\gamma)\mu^\top + \gamma d_\mu^\pi P^\pi = (1-\gamma)\mu^\top (I-\gamma P^\pi)^{-1}, \tag{23}$$

   and $d_\mu^\pi \mathbf{1} = 1$.

4. *The μ-weighted value satisfies*

$$V_i^\pi(\mu) = \mu^\top V_i^\pi = d_\mu^\pi r_i^\pi. \tag{24}$$

Proof. Because P^π is row-stochastic, $\|P^\pi\|_\infty = 1$, hence $\|\gamma P^\pi\|_\infty \le \gamma < 1$. Therefore the Neumann series

$$(I-\gamma P^\pi)^{-1} = \sum_{t=0}^{\infty}(\gamma P^\pi)^t$$

converges, which proves invertibility.

For (21), conditioning on the first step under π gives

$$V_i^\pi(s) = \sum_{a \in A(s)} \pi(a \mid s)\Big((1-\gamma) r_i(s,a) + \gamma \sum_{s'} P(s' \mid s,a) V_i^\pi(s')\Big),$$

which is exactly

$$V_i^\pi = (1-\gamma) r_i^\pi + \gamma P^\pi V_i^\pi.$$

Multiplying by $(I-\gamma P^\pi)^{-1}$ yields the closed form.

Identity (22) follows by substituting the definition of $Q_i^\pi(s,a)$ from (18) into the previous displayed equation.

For (23), note that

$$d_\mu^\pi = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t\mu^\top(P^\pi)^t = (1-\gamma)\mu^\top + \gamma(1-\gamma)\sum_{t=0}^{\infty}\gamma^t\mu^\top(P^\pi)^{t+1},$$

which is exactly

$$d_\mu^\pi = (1-\gamma)\mu^\top + \gamma d_\mu^\pi P^\pi.$$

The closed form again follows from the Neumann series. Since $P^\pi\mathbf{1}=\mathbf{1}$,

$$d_\mu^\pi\mathbf{1} = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t\mu^\top(P^\pi)^t\mathbf{1} = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t\mu^\top\mathbf{1} = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t = 1.$$

Finally,

$$d_\mu^\pi r_i^\pi = (1-\gamma)\mu^\top(I-\gamma P^\pi)^{-1}r_i^\pi = \mu^\top V_i^\pi = V_i^\pi(\mu),$$

where the middle equality uses (21). □

B.4. Best-Response Bellman Operators

For a fixed stationary opponent policy π_{-i}, define the Bellman best-response operator $T_i^{\pi_{-i}}:\mathbb{R}^S\to\mathbb{R}^S$ by

$$(T_i^{\pi_{-i}}v)(s) := \begin{cases} \max\limits_{a\in A(s)}\left\{(1-\gamma)r_i(s,a)+\gamma\sum\limits_{s'\in S}P(s'\mid s,a)\,v(s')\right\}, & s\in S_i,\\ \sum\limits_{a\in A(s)}\pi(a\mid s)\left\{(1-\gamma)r_i(s,a)+\gamma\sum\limits_{s'\in S}P(s'\mid s,a)\,v(s')\right\}, & s\in S_{-i}. \end{cases} \tag{25}$$

The operator $T_i^{\pi_{-i}}$ is the discounted optimality operator of the finite Markov decision problem faced by player i when the opponent commits to π_{-i}.

Lemma B.6 (Best-response operator)**.** *Fix $i\in\{1,2\}$ and a stationary opponent policy π_{-i}. Then $T_i^{\pi_{-i}}$ is monotone and a γ-contraction in the supremum norm. Consequently, it has a unique fixed point $W_i^\pi\in\mathbb{R}^S$, and this fixed point equals the best-response value function:*

$$W_i^\pi(s) = \sup_{\sigma_i} V_i^{(\sigma_i,\pi_{-i})}(s), \qquad s\in S. \tag{26}$$

Moreover, there exists an optimal deterministic stationary best response $\tau_i^\star\in\Pi_i(G)$ such that

$$W_i^\pi = V_i^{(\tau_i^\star,\pi_{-i})}.$$

Proof. Monotonicity is immediate from (25): if $u\le v$ componentwise, then every affine continuation term defining $T_i^{\pi_{-i}}u$ is bounded above by the corresponding term defining $T_i^{\pi_{-i}}v$, and both maximization and expectation preserve order.

For the contraction property, fix $u,v\in\mathbb{R}^S$. If $s\in S_i$, then

$$\begin{aligned}\left|(T_i^{\pi_{-i}}u)(s)-(T_i^{\pi_{-i}}v)(s)\right| &\le \max_{a\in A(s)}\gamma\left|\sum_{s'}P(s'\mid s,a)\big(u(s')-v(s')\big)\right|\\ &\le \gamma\|u-v\|_\infty.\end{aligned}$$

If $s\in S_{-i}$, the same estimate holds with the maximum replaced by the $\pi(\cdot\mid s)$-expectation. Taking the maximum over $s\in S$ yields

$$\|T_i^{\pi_{-i}}u - T_i^{\pi_{-i}}v\|_\infty \le \gamma\|u-v\|_\infty.$$

Hence $T_i^{\pi_{-i}}$ has a unique fixed point w by the Banach fixed-point theorem.

We next show that w dominates every unilateral deviation. For any behavioral strategy σ_i of player i, define

$$M_t := (1-\gamma)\sum_{\tau=0}^{t-1}\gamma^\tau r_i(s_\tau, a_\tau) + \gamma^t w(s_t).$$

We claim that under the play measure induced by (σ_i, π_{-i}), $(M_t)_{t\geq 0}$ is a supermartingale. Indeed, condition on an arbitrary history h_t ending at state s_t.

If $s_t \in S_i$, then by the fixed-point relation $w = T_i^{\pi_{-i}} w$,

$$w(s_t) = \max_{a\in A(s_t)} \left\{ (1-\gamma) r_i(s_t, a) + \gamma \sum_{s'} P(s' \mid s_t, a) w(s') \right\},$$

so for every action $a \in A(s_t)$,

$$w(s_t) \geq (1-\gamma) r_i(s_t, a) + \gamma \sum_{s'} P(s' \mid s_t, a) w(s').$$

Taking expectation over the (possibly history-dependent and randomized) action chosen by σ_i preserves the inequality.

If $s_t \in S_{-i}$, then by definition of $T_i^{\pi_{-i}}$ we have

$$w(s_t) = \sum_{a\in A(s_t)} \pi(a \mid s_t) \left\{ (1-\gamma) r_i(s_t, a) + \gamma \sum_{s'} P(s' \mid s_t, a) w(s') \right\},$$

which is equality under the opponent's stationary action distribution. In both cases,

$$\mathbb{E}[M_{t+1} \mid h_t] \leq M_t.$$

Taking expectations and iterating yields, for every horizon T,

$$w(s) \geq \mathbb{E}_s^{(\sigma_i, \pi_{-i})} \left[(1-\gamma) \sum_{t=0}^{T-1} \gamma^t r_i(s_t, a_t) + \gamma^T w(s_T) \right].$$

Since $w \in [0,1]^S$, the remainder term $\gamma^T w(s_T)$ vanishes as $T \to \infty$, so

$$w(s) \geq V_i^{(\sigma_i, \pi_{-i})}(s) \qquad \text{for all } s \in S.$$

Thus w upper-bounds the value of every unilateral deviation.

To show attainability by a stationary policy, for each state $s \in S_i$, choose an action

$$a^\star(s) \in \arg\max_{a\in A(s)} \left\{ (1-\gamma) r_i(s, a) + \gamma \sum_{s'} P(s' \mid s, a) w(s') \right\},$$

which is possible because $A(s)$ is finite. Let $\tau_i^\star$ be the deterministic stationary policy that plays $a^\star(s)$ at state s. Then

$$w = T_i^{(\tau_i^\star, \pi_{-i})} w,$$

where $T_i^{(\tau_i^\star, \pi_{-i})}$ is the linear Bellman evaluation operator for the fixed stationary profile $(\tau_i^\star, \pi_{-i})$. Since that operator is also a γ-contraction, its fixed point is unique and must equal the value vector $V_i^{(\tau_i^\star, \pi_{-i})}$. Hence

$$w = V_i^{(\tau_i^\star, \pi_{-i})}.$$

Combining the upper-bound argument with attainability shows that w is exactly the best-response value function. □

B.5. Bellman Characterization of Exact and Approximate Equilibrium

Proposition B.7 (Exact stationary Nash as Bellman complementarity)**.** *Let $\pi \in \Pi(G)$ be a stationary profile. The following are equivalent.*

1. *π is a statewise stationary Nash equilibrium.*
2. *For each player $i \in \{1,2\}$,*
$$T_i^{\pi_{-i}} V_i^\pi = V_i^\pi.$$
3. *For each player $i \in \{1,2\}$ and each state $s \in S_i$,*
$$V_i^\pi(s) = \max_{a \in A(s)} Q_i^\pi(s,a). \tag{27}$$
4. *For each player $i \in \{1,2\}$, each state $s \in S_i$, and each action $a \in A(s)$,*
$$\Delta_i^\pi(s,a) \le 0. \tag{28}$$

In particular, exact stationary Nash is equivalent to the vanishing of all positive one-step unilateral deviation gains.

Proof. $(1) \Leftrightarrow (2)$: by Definition B.3 with $\varepsilon = 0$, π is a statewise stationary Nash equilibrium iff $V_i^\pi = W_i^\pi$ for each player i. By Lemma B.6, W_i^π is the unique fixed point of $T_i^{\pi_{-i}}$. Hence (1) and (2) are equivalent.

$(2) \Leftrightarrow (3)$: for $s \in S_i$, the definition of $T_i^{\pi_{-i}}$ gives

$$(T_i^{\pi_{-i}} V_i^\pi)(s) = \max_{a \in A(s)} Q_i^\pi(s,a).$$

For $s \in S_{-i}$, the same operator equals

$$(T_i^{\pi_{-i}} V_i^\pi)(s) = \sum_{a \in A(s)} \pi(a \mid s)\, Q_i^\pi(s,a) = V_i^\pi(s)$$

by (22). Therefore the fixed-point equation $T_i^{\pi_{-i}} V_i^\pi = V_i^\pi$ is equivalent to (27) on the states controlled by player i.

$(3) \Leftrightarrow (4)$: by definition,

$$\Delta_i^\pi(s,a) = Q_i^\pi(s,a) - V_i^\pi(s).$$

Thus (27) is equivalent to saying that $Q_i^\pi(s,a) \le V_i^\pi(s)$ for all $a \in A(s)$, i.e., (28). □

The exact characterization above does *not* extend verbatim to ε-equilibria with the same ε, because Bellman errors can accumulate along future revisits. The next proposition records the precise comparison.

Proposition B.8 (Bellman residual versus approximate equilibrium)**.** *For each stationary profile $\pi \in \Pi(G)$ and player $i \in \{1,2\}$, define the Bellman residual*

$$\mathrm{res}_i(\pi) := \left\| T_i^{\pi_{-i}} V_i^\pi - V_i^\pi \right\|_\infty. \tag{29}$$

Then

$$\mathrm{res}_i(\pi) = \max_{s \in S_i} \max_{a \in A(s)} \Delta_i^\pi(s,a)_+, \tag{30}$$

where $x_+ = \max\{x, 0\}$. Moreover,

$$0 \;\le\; W_i^\pi - V_i^\pi \;\le\; \frac{\mathrm{res}_i(\pi)}{1-\gamma}\,\mathbf{1} \qquad \textit{componentwise.} \tag{31}$$

Consequently:

1. *if $\mathrm{res}_i(\pi) \le (1-\gamma)\varepsilon$ for both players $i \in \{1,2\}$, then π is a statewise stationary ε-Nash equilibrium;*

2. *if π is a statewise stationary ε-Nash equilibrium, then* $\mathrm{res}_i(\pi) \le \varepsilon$ *for both players* $i \in \{1,2\}$.

Proof. Fix a player i. For $s \in S_i$,

$$(T_i^{\pi_{-i}} V_i^\pi)(s) - V_i^\pi(s) = \max_{a \in A(s)} Q_i^\pi(s,a) - V_i^\pi(s) = \max_{a \in A(s)} \Delta_i^\pi(s,a).$$

For $s \in S_{-i}$, Proposition B.5 gives

$$(T_i^{\pi_{-i}} V_i^\pi)(s) = V_i^\pi(s).$$

Hence

$$\mathrm{res}_i(\pi) = \max_{s \in S_i} \max_{a \in A(s)} \Delta_i^\pi(s,a).$$

Since $V_i^\pi(s)$ is a convex combination of $\{Q_i^\pi(s,a) : a \in A(s)\}$, the quantity above is always nonnegative, which yields (30).

Next, because π_i is itself an admissible unilateral strategy,

$$W_i^\pi \ge V_i^\pi \qquad \text{componentwise.}$$

Using the fixed-point identity $W_i^\pi = T_i^{\pi_{-i}} W_i^\pi$, we obtain

$$\begin{aligned} \left\|W_i^\pi - V_i^\pi\right\|_\infty &= \left\|T_i^{\pi_{-i}} W_i^\pi - V_i^\pi\right\|_\infty \\ &\le \left\|T_i^{\pi_{-i}} W_i^\pi - T_i^{\pi_{-i}} V_i^\pi\right\|_\infty + \left\|T_i^{\pi_{-i}} V_i^\pi - V_i^\pi\right\|_\infty \\ &\le \gamma \left\|W_i^\pi - V_i^\pi\right\|_\infty + \mathrm{res}_i(\pi), \end{aligned}$$

where the last step uses Lemma B.6. Rearranging gives

$$\left\|W_i^\pi - V_i^\pi\right\|_\infty \le \frac{\mathrm{res}_i(\pi)}{1-\gamma},$$

which, combined with $W_i^\pi \ge V_i^\pi$, proves (31).

If $\mathrm{res}_i(\pi) \le (1-\gamma)\varepsilon$ for both players, then (31) gives

$$W_i^\pi(s) - V_i^\pi(s) \le \varepsilon \qquad \forall s \in S,\ i \in \{1,2\},$$

which is exactly the stationary ε-Nash condition.

Conversely, if π is a stationary ε-Nash equilibrium, then

$$W_i^\pi \le V_i^\pi + \varepsilon \mathbf{1}.$$

By monotonicity of $T_i^{\pi_{-i}}$,

$$T_i^{\pi_{-i}} V_i^\pi \le T_i^{\pi_{-i}} W_i^\pi = W_i^\pi \le V_i^\pi + \varepsilon \mathbf{1},$$

hence

$$\mathrm{res}_i(\pi) = \left\|T_i^{\pi_{-i}} V_i^\pi - V_i^\pi\right\|_\infty \le \varepsilon.$$

□

B.6. Bellman Jet

Definition B.9 (Bellman jet)**.** For a stationary profile $\pi \in \Pi(G)$, the *Bellman jet* of π in game G is the tuple

$$J_G(\pi) := \Big(\pi,\ V_1^\pi, V_2^\pi,\ Q_1^\pi, Q_2^\pi,\ \Delta_1^\pi, \Delta_2^\pi,\ d_\mu^\pi\Big).$$

Here $V_i^\pi \in \mathbb{R}^S$, $Q_i^\pi \in \mathbb{R}^{\{(s,a): a \in A(s)\}}$, $\Delta_i^\pi \in \mathbb{R}^{\{(s,a): a \in A(s)\}}$, and $d_\mu^\pi \in \mathbb{R}^{1 \times S}$.

Remark B.10. The Bellman jet packages precisely the local continuation information that our later impossibility theorem allows a learning dynamic to inspect: the current policy, exact value vectors, exact one-step continuation values, exact deviation gains, and exact discounted occupancy weights.

C. Exact Bellman-Jet Computability

This appendix proves that the Bellman jet introduced in Definition B.9 is exactly computable in polynomial time in the standard Turing bit model when the game instance and the stationary policy profile are rationally encoded. The only nontrivial ingredient is exact linear algebra over the rationals; for this we use standard fraction-free elimination ideas (Bareiss, 1968) and polynomial-time exact integer linear algebra results (Kannan & Bachem, 1979).

C.1. Bit Model and Encoding Conventions

Throughout this appendix, all complexity statements are in the standard Turing bit model. For a rational number $q = a/b \in \mathbb{Q}$ written in lowest terms with $b > 0$, define its binary encoding length by

$$\langle q \rangle := \lceil \log_2(|a| + 1) \rceil + \lceil \log_2(b + 1) \rceil .$$

For a rational vector or matrix, $\langle \cdot \rangle$ denotes the total encoding length obtained by summing the encoding lengths of all entries, up to the standard $O(1)$-factor delimiter overhead. For a game instance G and a rational stationary profile π, $\langle G \rangle$ and $\langle \pi \rangle$ denote their full binary encoding lengths.

Fix a discounted perfect-information stochastic game G as in Definition B.1, and a rational stationary profile $\pi \in \Pi(G)$. Let

$$n := |S|, \qquad m := \sum_{s \in S} |A(s)|, \qquad N := \langle G \rangle + \langle \pi \rangle.$$

Since the input explicitly lists the state set, action sets, transition table, reward table, discount factor, initial distribution, and stationary action probabilities, both n and m are bounded above by N up to absolute constants.

We also define

$$\tau := \max\Big\{ \langle \gamma \rangle,\ \langle \mu(s) \rangle,\ \langle r_i(s,a) \rangle,\ \langle P(s' \mid s,a) \rangle,\ \langle \pi(a \mid s) \rangle \Big\},$$

where the maximum ranges over all indices for which the displayed quantities are defined. Again, $\tau \le N$.

C.2. Auxiliary Bit-Growth Lemmas

Lemma C.1 (Bit growth under exact rational arithmetic)**.** *Let $q_1, \dots, q_k \in \mathbb{Q}$ be rationals with $\langle q_j \rangle \le L$ for all $j \in [k]$. Then the quantities*

$$\sum_{j=1}^{k} q_j \qquad \text{and} \qquad \prod_{j=1}^{k} q_j$$

can be computed exactly in time polynomial in k and L, and each has encoding length $O(kL + k \log k)$.

Proof. Write $q_j = a_j/b_j$ in lowest terms with $b_j > 0$. Since $\langle q_j \rangle \le L$, we have

$$|a_j| \le 2^L - 1, \qquad 1 \le b_j \le 2^L - 1.$$

For the product,

$$\prod_{j=1}^{k} q_j = \frac{\prod_{j=1}^{k} a_j}{\prod_{j=1}^{k} b_j}.$$

Hence both numerator and denominator have binary length at most kL, so the result has encoding length $O(kL)$. Exact computation uses $k - 1$ integer multiplications in numerator and denominator, hence polynomial time.

For the sum, use the common denominator

$$B := \prod_{j=1}^{k} b_j.$$

Then

$$\sum_{j=1}^{k} q_j = \frac{A}{B}, \qquad A := \sum_{j=1}^{k} a_j \prod_{\ell \ne j} b_\ell.$$

The denominator B has binary length at most kL. Each summand in A has binary length at most kL, and adding k integers of bit length $O(kL)$ yields an integer of bit length $O(kL + \log k)$. Therefore the sum has encoding length $O(kL + k \log k)$. Exact computation again requires only polynomially many exact integer additions and multiplications, followed optionally by Euclidean reduction of the final fraction. □

Lemma C.2 (Exact solution of rational linear systems)**.** *Let $A \in \mathbb{Q}^{n\times n}$ be invertible and let $b \in \mathbb{Q}^n$. Assume every entry of A and b has encoding length at most L. Then the unique solution $x = A^{-1}b$ can be computed exactly in time polynomial in n and L, and every coordinate of x has encoding length polynomial in n and L. More concretely,*

$$\max_{j\in[n]} \langle x_j \rangle = O(n^3 L + n^2 \log n).$$

Proof. Choose a positive integer D equal to the product of the denominators of all entries of A and b. Since there are $n^2 + n$ such entries and each denominator has bit length at most L, we have

$$\langle D \rangle \le (n^2 + n)L.$$

Define

$$\widehat{A} := DA \in \mathbb{Z}^{n\times n}, \qquad \widehat{b} := Db \in \mathbb{Z}^n.$$

Then $\widehat{A}$ is invertible over $\mathbb{Q}$, and

$$Ax = b \quad \Longleftrightarrow \quad \widehat{A}x = \widehat{b}.$$

Every entry of $\widehat{A}$ and $\widehat{b}$ has encoding length $O(n^2 L)$.

To bound the output size, write x_j using Cramer's rule:

$$x_j = \frac{\det(\widehat{A}^{(j)})}{\det(\widehat{A})},$$

where $\widehat{A}^{(j)}$ is obtained from $\widehat{A}$ by replacing its j-th column by $\widehat{b}$. Let M be the maximum absolute value of an entry of $\widehat{A}$ or $\widehat{b}$. Then $M < 2^{cn^2 L}$ for some absolute constant c. By Hadamard's inequality,

$$|\det(\widehat{A})| \le n^{n/2} M^n, \qquad |\det(\widehat{A}^{(j)})| \le n^{n/2} M^n.$$

Hence both numerator and denominator in Cramer's rule have binary length

$$O\big(n \log n + n \log M\big) = O(n^3 L + n^2 \log n).$$

Therefore each coordinate of x has encoding length polynomial in n and L.

For exact computation, one may solve the integer system $\widehat{A}x = \widehat{b}$ using fraction-free Gaussian elimination with pivoting, as in Bareiss (1968); see also Kannan & Bachem (1979) for polynomial-time exact integer linear algebra with explicit control of intermediate bit growth. Since all matrix entries have polynomially bounded bit-length and the matrix dimension is n, the exact running time is polynomial in n and L. □

C.3. Exact Computation of the Induced Bellman Data

Recall from Equations (15) and (16) that

$$P^\pi_{ss'} = \sum_{a\in A(s)} \pi(a \mid s) P(s' \mid s, a), \qquad r^\pi_i(s) = \sum_{a\in A(s)} \pi(a \mid s) r_i(s, a).$$

Lemma C.3 (Exact computation of P^π and r^π_i)**.** *The induced transition matrix P^π and reward vectors r^π_i, $i \in \{1, 2\}$, can be computed exactly in time polynomial in N. Moreover, every entry of P^π and of each r^π_i has encoding length polynomial in N.*

Proof. Fix $s, s' \in S$. The quantity $P^\pi_{ss'}$ is a sum of at most $|A(s)| \le m$ products of rational numbers. Each factor $\pi(a \mid s)$ and $P(s' \mid s, a)$ has encoding length at most τ, so each product has encoding length $O(\tau)$. By Lemma C.1, the sum over $a \in A(s)$ has encoding length $O(m\tau + m \log m)$, hence polynomial in N, and is exactly computable in polynomial time. The argument for $r^\pi_i(s)$ is identical. □

For convenience, define

$$A^\pi := I - \gamma P^\pi.$$

Proposition C.4 (Exact computation of value vectors and occupancy)**.** *For each player $i \in \{1,2\}$, the value vector V_i^π is exactly computable in time polynomial in N, and every entry of V_i^π has encoding length polynomial in N. The same holds for the discounted occupancy vector d_μ^π.*

Proof. By Lemma C.3, the matrix P^π and vectors r_i^π have entries of polynomial encoding length and are exactly computable in polynomial time. Therefore every entry of $A^\pi = I - \gamma P^\pi$ also has polynomial encoding length and is exactly computable in polynomial time.

By Proposition B.5, the matrix A^π is invertible and V_i^π is the unique solution to

$$A^\pi V_i^\pi = (1-\gamma) r_i^\pi. \tag{32}$$

All coefficients in (32) have encoding length polynomial in N. Hence Lemma C.2 applies and yields exact polynomial-time computability of V_i^π, together with a polynomial bound on the encoding length of each coordinate.

For the occupancy vector, Proposition B.5 gives

$$d_\mu^\pi = (1-\gamma)\mu^\top (I - \gamma P^\pi)^{-1}.$$

Equivalently, the transpose $x := (d_\mu^\pi)^\top$ is the unique solution of

$$(A^\pi)^\top x = (1-\gamma)\mu. \tag{33}$$

Again all coefficients have polynomial encoding length, so Lemma C.2 applies and yields exact polynomial-time computability of d_μ^π, with polynomial output bit complexity. □

Lemma C.5 (Exact computation of Q_i^π and Δ_i^π)**.** *For each player $i \in \{1,2\}$, the one-step continuation table Q_i^π and the deviation-gain table Δ_i^π are exactly computable in time polynomial in N, and every entry of both tables has encoding length polynomial in N.*

Proof. Fix $i \in \{1,2\}$, $s \in S$, and $a \in A(s)$. By Equation (18),

$$Q_i^\pi(s,a) = (1-\gamma) r_i(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a) V_i^\pi(s').$$

The first term has encoding length at most $O(\tau)$. For the second term, each factor $P(s' \mid s,a)$ and $V_i^\pi(s')$ has polynomial encoding length by Proposition C.4, so each product has polynomial encoding length, and the sum over at most $n \le N$ states remains of polynomial encoding length by Lemma C.1. Hence $Q_i^\pi(s,a)$ is exactly computable in polynomial time with polynomial output encoding length.

Finally,

$$\Delta_i^\pi(s,a) = Q_i^\pi(s,a) - V_i^\pi(s),$$

so exact polynomial-time computability and polynomial encoding-length bounds for $\Delta_i^\pi(s,a)$ follow immediately from exact rational subtraction. □

C.4. Main Exact-Computation Theorem

Theorem C.6 (Exact Bellman-jet computability)**.** *There exists a polynomial p such that for every rational discounted perfect-information stochastic game G and every rational stationary profile $\pi \in \Pi(G)$, the Bellman jet*

$$J_G(\pi) = \Big(\pi,\ V_1^\pi, V_2^\pi,\ Q_1^\pi, Q_2^\pi,\ \Delta_1^\pi, \Delta_2^\pi,\ d_\mu^\pi\Big)$$

can be computed exactly in time at most $p(\langle G\rangle + \langle \pi\rangle)$. Moreover, every coordinate appearing in $J_G(\pi)$ has encoding length at most $p(\langle G\rangle + \langle \pi\rangle)$.

Proof. The policy profile π is part of the input. By Lemma C.3, we can exactly compute P^π and both reward vectors r_1^π, r_2^π in polynomial time, with polynomially bounded encoding lengths. By Proposition C.4, we can then exactly compute the value vectors V_1^π, V_2^π and the discounted occupancy vector d_μ^π, again in polynomial time and with polynomially bounded encoding lengths. Finally, by Lemma C.5, we can exactly compute the tables $Q_1^\pi, Q_2^\pi, \Delta_1^\pi, \Delta_2^\pi$, with the same polynomial bounds.

Concatenating these polynomial-time procedures yields an exact polynomial-time algorithm for the entire Bellman jet. Since the Bellman jet contains only a polynomial number of rational entries, each of polynomial encoding length, its total encoding length is also polynomial in $N = \langle G \rangle + \langle \pi \rangle$. □

Corollary C.7 (Bounded-bit stationary policy grids)**.** *Let $b : \mathbb{N} \to \mathbb{N}$ be any function, and suppose π is a stationary profile whose every probability entry has encoding length at most $b(\langle G \rangle)$. Then the Bellman jet $J_G(\pi)$ is exactly computable in time polynomial in $\langle G \rangle + b(\langle G \rangle)$, and its total encoding length is polynomial in the same quantity.*

Proof. Immediate from Theorem C.6, since $\langle \pi \rangle$ is polynomially bounded by the number of policy entries times the entry-wise bound $b(\langle G \rangle)$, and the number of policy entries is itself polynomial in $\langle G \rangle$. □

D. Rational Encodings, Finite Search Domain, and Strict Descent Termination

This appendix fixes the finite search domain on which our later local-search reduction operates. The key point is that once one restricts attention to stationary policies whose action probabilities have bounded rational encoding length, the space of policies becomes a finite, canonically encoded set. This allows Bellman-local dynamics to be interpreted as discrete improvement maps on a finite search domain.

Throughout this appendix, G is a fixed discounted perfect-information stochastic game as in Definition B.1, and $b \in \mathbb{N}$ is a bit budget with $b \geq 1$. We retain the notation

$$m := \sum_{s \in S} |A(s)|$$

for the total number of state-action coordinates appearing in a stationary policy profile.

D.1. Bounded-Bit Rational Probabilities

We begin by fixing a canonical finite set of rational probabilities of bounded bit complexity.

Definition D.1 (b-bit rational probabilities)**.** For $b \geq 1$, define

$$\mathbb{Q}_{[0,1]}^{(b)} := \left\{ \frac{u}{v} \in \mathbb{Q} : 0 \leq u \leq v < 2^b,\ 1 \leq v,\ \gcd(u, v) = 1 \right\}.$$

Equivalently, $\mathbb{Q}_{[0,1]}^{(b)}$ is the set of all rational numbers in $[0, 1]$ whose unique reduced numerator and denominator are both representable using at most b binary digits.

Remark D.2. The numbers 0 and 1 always belong to $\mathbb{Q}_{[0,1]}^{(b)}$ for every $b \geq 1$, via the unique reduced representations

$$0 = \frac{0}{1}, \qquad 1 = \frac{1}{1}.$$

Lemma D.3 (Finiteness of the bounded-bit probability set)**.** *For every $b \geq 1$,*

$$1 \leq \left| \mathbb{Q}_{[0,1]}^{(b)} \right| \leq 2^{2b}.$$

Moreover, every $q \in \mathbb{Q}_{[0,1]}^{(b)}$ has a unique reduced representation $q = u/v$ satisfying the conditions in Definition D.1.

Proof. The lower bound is immediate from $\{0, 1\} \subseteq \mathbb{Q}_{[0,1]}^{(b)}$. For the upper bound, the pair (u, v) ranges over at most

$$2^b \cdot (2^b - 1) < 2^{2b}$$

possibilities, since $u \in \{0, \dots, 2^b - 1\}$ and $v \in \{1, \dots, 2^b - 1\}$. The set $\mathbb{Q}^{(b)}_{[0,1]}$ is obtained by restricting to those pairs with $u \le v$ and $\gcd(u, v) = 1$, hence it is finite and has cardinality at most 2^{2b}.

Uniqueness follows from uniqueness of reduced rational representations: if

$$\frac{u}{v} = \frac{u'}{v'}$$

with $u, v, u', v' \in \mathbb{Z}_{\ge 0}$, $v, v' > 0$, and $\gcd(u, v) = \gcd(u', v') = 1$, then $u = u'$ and $v = v'$. □

D.2. Bounded-Bit Stationary Policy Grid

We now discretize the stationary policy space.

Definition D.4 (b-bit stationary policy grid)**.** Let G be fixed and $b \ge 1$. Define

$$\Pi_b(G) := \left\{ \pi \in \Pi(G) : \pi(a \mid s) \in \mathbb{Q}^{(b)}_{[0,1]} \text{ for every } s \in S,\ a \in A(s) \right\}.$$

Thus $\Pi_b(G)$ is the set of stationary policy profiles whose every action probability lies in the finite set $\mathbb{Q}^{(b)}_{[0,1]}$.

Remark D.5. The set $\Pi_b(G)$ is not a uniform denominator lattice. Different action probabilities at the same state are allowed to have different reduced denominators, provided each coordinate belongs to $\mathbb{Q}^{(b)}_{[0,1]}$ and the resulting action probabilities sum exactly to 1 at each state.

To obtain a canonical binary encoding, fix once and for all a total order on the set of action coordinates

$$\mathcal{I}(G) := \{(s, a) : s \in S,\ a \in A(s)\},$$

for example the order induced by the input listing of states and, within each state, the input listing of available actions. Enumerate

$$\mathcal{I}(G) = \{(s_1, a_1), \dots, (s_m, a_m)\},$$

where $m = \sum_{s \in S} |A(s)|$.

Definition D.6 (Canonical b-bit encoding)**.** For $\pi \in \Pi_b(G)$, write each action probability in unique reduced form

$$\pi(a_j \mid s_j) = \frac{u_j}{v_j}, \qquad 0 \le u_j \le v_j < 2^b, \quad 1 \le v_j, \quad \gcd(u_j, v_j) = 1.$$

The *canonical b-bit encoding* of π, denoted $\mathrm{enc}_{G,b}(\pi)$, is the binary string obtained by concatenating, in the fixed order $j = 1, \dots, m$, the b-bit zero-padded binary representations of u_j and v_j. Its total length is

$$\ell_b(G) := 2bm. \tag{34}$$

Define the ambient cube

$$\mathcal{X}_b(G) := \{0, 1\}^{\ell_b(G)}.$$

Definition D.7 (Validity and decoding)**.** Given $x \in \mathcal{X}_b(G)$, parse x into m consecutive pairs of b-bit blocks

$$(u_1, v_1), \dots, (u_m, v_m),$$

interpreted as nonnegative integers in $\{0, \dots, 2^b - 1\}$. We call x *valid* if all of the following hold:

1. $1 \le v_j$ for every $j \in [m]$;
2. $0 \le u_j \le v_j$ for every $j \in [m]$;
3. $\gcd(u_j, v_j) = 1$ for every $j \in [m]$;
4. for every state $s \in S$,

$$\sum_{a \in A(s)} \frac{u_{(s,a)}}{v_{(s,a)}} = 1 \qquad \text{exactly in } \mathbb{Q}. \tag{35}$$

If x is valid, we define $\mathrm{dec}_{G,b}(x) \in \Pi_b(G)$ by

$$\mathrm{dec}_{G,b}(x)(a_j \mid s_j) := \frac{u_j}{v_j}.$$

If x is invalid, we set

$$\mathrm{dec}_{G,b}(x) := \bot,$$

where $\bot$ is a formal symbol not belonging to $\Pi_b(G)$. We denote the set of valid encodings by

$$\mathcal{V}_b(G) := \{x \in \mathcal{X}_b(G) : \mathrm{dec}_{G,b}(x) \neq \bot\}.$$

Lemma D.8 (Canonical encoding is a bijection)**.** *For every fixed G and $b \geq 1$, the map*

$$\mathrm{enc}_{G,b} : \Pi_b(G) \to \mathcal{V}_b(G)$$

is a bijection, with inverse $\mathrm{dec}_{G,b}|_{\mathcal{V}_b(G)}$. Moreover, the following tasks are all executable in time polynomial in $\langle G \rangle + b$:

1. *compute $\mathrm{enc}_{G,b}(\pi)$ from $\pi \in \Pi_b(G)$;*
2. *test whether $x \in \mathcal{X}_b(G)$ is valid;*
3. *compute $\mathrm{dec}_{G,b}(x)$ for valid x.*

Proof. We first prove that $\mathrm{enc}_{G,b}$ maps $\Pi_b(G)$ into $\mathcal{V}_b(G)$. Let $\pi \in \Pi_b(G)$. By Definition D.4, every coordinate $\pi(a_j \mid s_j)$ lies in $\mathbb{Q}^{(b)}_{[0,1]}$, so it has a unique reduced representation u_j/v_j satisfying the constraints in Definition D.1. Because $\pi \in \Pi(G)$, the exact simplex condition (35) holds at every state. Hence $\mathrm{enc}_{G,b}(\pi)$ is valid, and

$$\mathrm{dec}_{G,b}(\mathrm{enc}_{G,b}(\pi)) = \pi.$$

Conversely, if $x \in \mathcal{V}_b(G)$, then the parsed fractions satisfy all the defining constraints, including the exact simplex condition at each state, so $\mathrm{dec}_{G,b}(x) \in \Pi_b(G)$. By uniqueness of reduced representations (Lemma D.3), re-encoding recovers the same bit string:

$$\mathrm{enc}_{G,b}(\mathrm{dec}_{G,b}(x)) = x.$$

Thus $\mathrm{enc}_{G,b}$ is a bijection with inverse $\mathrm{dec}_{G,b}|_{\mathcal{V}_b(G)}$.

For the algorithmic statements, encoding consists of reading off the reduced numerator-denominator pair in each coordinate and zero-padding each to b bits. Validity checking requires: (i) verifying the positivity and order constraints on each pair, (ii) checking $\gcd(u_j, v_j) = 1$ via the Euclidean algorithm, and (iii) checking the exact simplex constraint (35) for each state. Task (iii) can be done by repeated exact rational addition, with polynomial bit growth guaranteed by Lemma C.1. Since the number of coordinates is $m \leq \langle G \rangle$ up to constants, the overall running time is polynomial in $\langle G \rangle + b$. Decoding valid encodings is immediate once the parsed fractions have been certified. □

Proposition D.9 (Finite search-domain bounds)**.** *For every fixed G and $b \geq 1$,*

$$1 \leq |\Pi_b(G)| = |\mathcal{V}_b(G)| \leq |\mathcal{X}_b(G)| = 2^{\ell_b(G)}.$$

In particular,

$$|\Pi_b(G)| \leq 2^{2bm}.$$

If b is polynomially bounded in $\langle G \rangle$, then the encoding length $\ell_b(G)$ is polynomially bounded in $\langle G \rangle$.

Proof. By Lemma D.8,

$$|\Pi_b(G)| = |\mathcal{V}_b(G)|.$$

Since $\mathcal{V}_b(G) \subseteq \mathcal{X}_b(G)$ and $|\mathcal{X}_b(G)| = 2^{\ell_b(G)}$, the upper bound is immediate.

For nonemptiness, define a deterministic stationary policy by selecting one action in each state and assigning it probability 1, with all remaining actions receiving probability 0. By Remark D.2, both 0 and 1 belong to $\mathbb{Q}^{(b)}_{[0,1]}$, so such a policy lies in $\Pi_b(G)$. The final claim follows from $\ell_b(G) = 2bm$ and the fact that m is explicitly represented in the game instance and hence polynomial in $\langle G \rangle$. □

D.3. Canonical Dummy Policy

For later use on invalid encodings, we fix a canonical valid policy profile that can serve as a sink or default decoding target.

Definition D.10 (Canonical deterministic dummy policy)**.** For each state $s \in S$, let $a^\circ(s) \in A(s)$ denote the first action in the fixed input order on $A(s)$. Define $\pi^\circ \in \Pi(G)$ by

$$\pi^\circ(a \mid s) := \begin{cases} 1, & a = a^\circ(s), \\ 0, & a \neq a^\circ(s). \end{cases}$$

For $b \geq 1$, let

$$x^\circ_{G,b} := \mathrm{enc}_{G,b}(\pi^\circ) \in \mathcal{V}_b(G)$$

denote its canonical encoding.

Lemma D.11 (Dummy policy is always available)**.** *For every discounted perfect-information stochastic game G and every $b \geq 1$,*

$$\pi^\circ \in \Pi_b(G), \qquad x^\circ_{G,b} \in \mathcal{V}_b(G).$$

Moreover, $x^\circ_{G,b}$ is computable in time polynomial in $\langle G \rangle + b$.

Proof. By construction, each coordinate of π° is either 0 or 1, both of which lie in $\mathbb{Q}^{(b)}_{[0,1]}$ by Remark D.2. The exact simplex constraints are also satisfied statewise because exactly one action at each state receives probability 1. Hence $\pi^\circ \in \Pi_b(G)$, and therefore $x^\circ_{G,b} = \mathrm{enc}_{G,b}(\pi^\circ) \in \mathcal{V}_b(G)$ by Lemma D.8. Polynomial-time computability is immediate from the definition. □

D.4. Strict Descent on a Finite Policy Grid

We next record the finite-improvement principle that underlies our later barrier theorem. The target set is kept abstract in this appendix because, in the main paper, it will later be instantiated by the set of stationary Nash equilibria inside the bounded-bit grid.

Definition D.12 (Strict descent pair relative to a target set)**.** Fix G, $b \geq 1$, and a subset $\mathcal{E} \subseteq \Pi_b(G)$. A pair of maps

$$F : \Pi_b(G) \to \Pi_b(G), \qquad L : \Pi_b(G) \to \mathbb{Q}$$

is called a *strict descent pair relative to $\mathcal{E}$* if:

1. $F(\pi) = \pi$ for every $\pi \in \mathcal{E}$;
2. $L(F(\pi)) < L(\pi)$ for every $\pi \in \Pi_b(G) \setminus \mathcal{E}$.

For an initial profile $\pi^{(0)} \in \Pi_b(G)$, we write

$$\pi^{(t+1)} := F(\pi^{(t)}), \qquad t \geq 0,$$

for the forward orbit of F.

Proposition D.13 (Strict descent implies finite termination)**.** *Let (F, L) be a strict descent pair relative to $\mathcal{E} \subseteq \Pi_b(G)$. Then:*

1. *every fixed point of F belongs to $\mathcal{E}$;*
2. *for every initial profile $\pi^{(0)} \in \Pi_b(G)$, the forward orbit enters $\mathcal{E}$ after at most $|\Pi_b(G)| - 1$ updates;*
3. *once the orbit enters $\mathcal{E}$, it remains there forever.*

In particular, the existence of a strict descent pair implies $\mathcal{E} \neq \varnothing$.

Proof. For item 1, suppose $F(\pi) = \pi$. If $\pi \notin \mathcal{E}$, then strict descent would imply

$$L(\pi) = L(F(\pi)) < L(\pi),$$

a contradiction. Hence every fixed point of F lies in $\mathcal{E}$.

For item 2, consider an orbit $(\pi^{(t)})_{t\geq 0}$. If $\pi^{(t)} \notin \mathcal{E}$, then by Definition D.12,

$$L(\pi^{(t+1)}) < L(\pi^{(t)}).$$

Therefore as long as the orbit remains outside $\mathcal{E}$, the values of L strictly decrease, and in particular the points $\pi^{(t)}$ are all distinct. Indeed, if $\pi^{(t')} = \pi^{(t)}$ for some $0 \leq t' < t$ while all intermediate points lie outside $\mathcal{E}$, then

$$L(\pi^{(t')}) > L(\pi^{(t'+1)}) > \cdots > L(\pi^{(t)}) = L(\pi^{(t')}),$$

a contradiction.

Since $\Pi_b(G)$ is finite, an orbit can visit at most $|\Pi_b(G)|$ distinct points. Hence it cannot remain outside $\mathcal{E}$ for $|\Pi_b(G)|$ consecutive iterates. Therefore there exists

$$t \leq |\Pi_b(G)| - 1$$

such that $\pi^{(t)} \in \mathcal{E}$.

For item 3, if $\pi^{(t)} \in \mathcal{E}$, then by item 1 of Definition D.12,

$$\pi^{(t+1)} = F(\pi^{(t)}) = \pi^{(t)},$$

and so the orbit stays constant from that time onward.

Finally, item 2 shows that every orbit reaches $\mathcal{E}$, so in particular $\mathcal{E}$ must be nonempty. □

Corollary D.14 (No cycles outside the target set)**.** *Under the assumptions of Proposition D.13, there is no directed cycle of F contained entirely in $\Pi_b(G) \setminus \mathcal{E}$.*

Proof. Any such cycle would force a strict chain

$$L(\pi_1) > L(\pi_2) > \cdots > L(\pi_k) > L(\pi_1),$$

which is impossible. □

D.5. Invalid Encodings and Dummy-Sink Handling

The local-search reduction in the main theorem will operate on the full cube $\mathcal{X}_b(G)$, not just on the valid subset $\mathcal{V}_b(G)$. The next definition and proposition isolate the generic argument showing that invalid encodings can be ruled out as local optima by redirecting them to a canonical valid sink.

Definition D.15 (Successor-objective pair on the full cube)**.** Let $\mathcal{X}$ be a finite set. A *successor-objective pair* on $\mathcal{X}$ is a pair of maps

$$S : \mathcal{X} \to \mathcal{X}, \qquad \Phi : \mathcal{X} \to \mathbb{Z}.$$

A point $x \in \mathcal{X}$ is called a *local optimum* of (S, Φ) if

$$\Phi(S(x)) \leq \Phi(x).$$

Equivalently, x admits no strict improvement along its designated successor.

Proposition D.16 (Invalid encodings cannot be local optima)**.** *Fix G and $b \geq 1$, and let $x^{\circ}_{G,b} \in \mathcal{V}_b(G)$ be the dummy encoding from Definition D.10. Suppose*

$$S : \mathcal{X}_b(G) \to \mathcal{X}_b(G), \qquad \Phi : \mathcal{X}_b(G) \to \mathbb{Z}$$

satisfy the following for every invalid string $x \in \mathcal{X}_b(G) \setminus \mathcal{V}_b(G)$:

1. $S(x) = x^{\circ}_{G,b}$;
2. $\Phi(x) < \Phi(x^{\circ}_{G,b})$.

Then every local optimum of (S, Φ) *belongs to* $\mathcal{V}_b(G)$.

Proof. Let $x \in \mathcal{X}_b(G) \setminus \mathcal{V}_b(G)$ be invalid. By assumption,

$$S(x) = x^{\circ}_{G,b} \qquad \text{and} \qquad \Phi(x^{\circ}_{G,b}) > \Phi(x).$$

Hence

$$\Phi(S(x)) > \Phi(x),$$

so x is not a local optimum in the sense of Definition D.15. Therefore every local optimum must lie in $\mathcal{V}_b(G)$. □

Remark D.17. Proposition D.16 is purely formal but crucial for the later reduction: it allows one to define local-search circuits on the entire hypercube $\mathcal{X}_b(G)$ while ensuring that all local optima are forced into the canonically encoded policy grid $\mathcal{V}_b(G)$.

E. Proof of the Main Barrier Theorem

This appendix proves the main barrier theorem. The proof has two steps. First, assuming the existence of a Bellman-local strict descent pair, we show that the bounded-bit exact stationary equilibrium problem belongs to PLS in the standard sense of polynomial local search (Johnson et al., 1988). Second, we combine this with the PPAD-completeness of exact stationary Nash equilibrium in two-player discounted perfect-information stochastic games (Hansen & Nie, 2025; Daskalakis et al., 2023) and the identity $\mathrm{CLS} = \mathrm{PPAD} \cap \mathrm{PLS}$ (Fearnley et al., 2022) to obtain the announced complexity-collapse corollary.

E.1. Problem Statement and Proof Strategy

Fix a polynomial bit-budget function $b : \mathbb{N} \to \mathbb{N}$, and recall the bounded-bit stationary policy grid $\Pi_b(G)$ and the exact grid-restricted equilibrium set

$$\mathrm{SNE}_b(G) = \mathrm{SNE}(G) \cap \Pi_b(G).$$

We first study the bounded-bit search problem

$$\mathsf{GridSNE}^b_{\mathrm{PI}}: \quad \text{input } G \in \mathcal{G}_{\mathrm{PI}}, \quad \text{output any } \pi \in \mathrm{SNE}_b(G).$$

Under the existence of a strict Bellman-Lyapunov descent pair, $\mathrm{SNE}_b(G) \neq \varnothing$ for every input G by Proposition D.13. Hence $\mathsf{GridSNE}^b_{\mathrm{PI}}$ is a total search problem under the hypothesis of the theorem.

The proof proceeds by constructing, for each game G, a discrete local-search instance on the full Boolean hypercube $\mathcal{X}_b(G)$ from Definition D.6. The designated successor of a valid encoding $x = \mathrm{enc}_{G,b}(\pi)$ will be the encoding of the next Bellman-local update $F_G(\pi)$. The delicate point is that the witness $L_G(\pi)$ is rational-valued, whereas the standard PLS formulation uses integer objectives. We therefore begin by showing that the witness values admit a polynomially bounded exact integerization that preserves all strict inequalities.

E.2. Polynomial Output Bounds for Bellman-Local Witnesses

Lemma E.1 (Polynomial bit bound for Bellman-local witness values)**.** *Let* $L = \{L_G\}_{G \in \mathcal{G}_{\mathrm{PI}}}$ *be a Bellman-local Lyapunov witness relative to the polynomial bit budget* b*. Then there exists a polynomial* q_L *such that for every* $G \in \mathcal{G}_{\mathrm{PI}}$ *and every* $\pi \in \Pi_b(G)$*,*

$$\langle L_G(\pi) \rangle \leq q_L(\langle G \rangle + b(\langle G \rangle)).$$

Proof. By Lemma 2.8, there exists a polynomial q_J such that for every $G \in \mathcal{G}_{\mathrm{PI}}$ and every $\pi \in \Pi_b(G)$, the Bellman jet $J_G(\pi)$ can be computed exactly in time at most $q_J(\langle G \rangle + b_G)$, and its full encoding length is at most $q_J(\langle G \rangle + b_G)$, where $b_G := b(\langle G \rangle)$.

Since L is Bellman-local, there exists a deterministic algorithm $\mathcal{A}_L$ and a polynomial t_L such that, on input a rational encoding of $J_G(\pi)$, the algorithm outputs $L_G(\pi)$ in time at most $t_L(|\operatorname{enc}(J_G(\pi))|)$. No Turing machine can output more bits than its running time. Therefore

$$\langle L_G(\pi)\rangle \le t_L\big(q_J(\langle G\rangle + b_G)\big).$$

Setting

$$q_L(n) := t_L(q_J(n))$$

proves the claim. □

For brevity, write

$$n_G := \langle G\rangle + b(\langle G\rangle), \qquad \beta_G := q_L(n_G).$$

E.3. Order-Preserving Integerization of Rational Witness Values

We now define a universal scaling scheme that converts witness values of encoding length at most β_G into nonnegative integers while preserving strict order.

Lemma E.2 (Separation of bounded-bit rationals)**.** *Let $B \ge 1$, and let $\alpha, \beta \in \mathbb{Q}$ satisfy*

$$\langle\alpha\rangle \le B, \qquad \langle\beta\rangle \le B, \qquad \alpha < \beta.$$

Then

$$\beta - \alpha > 2^{-2B}.$$

Proof. Write

$$\alpha = \frac{a}{c}, \qquad \beta = \frac{b}{d},$$

in lowest terms, with $c, d > 0$. Since $\langle\alpha\rangle \le B$ and $\langle\beta\rangle \le B$, we have

$$|a| < 2^B, \qquad c < 2^B, \qquad |b| < 2^B, \qquad d < 2^B.$$

Because $\alpha < \beta$,

$$\beta - \alpha = \frac{bc - ad}{cd},$$

where $bc - ad$ is a strictly positive integer. Therefore

$$\beta - \alpha \ge \frac{1}{cd} > 2^{-2B}.$$

□

Definition E.3 (Integerization parameters)**.** For each input game G, define

$$R_G := 2^{\beta_G+1}, \qquad C_G := 2^{2\beta_G+1}.$$

For every rational r with $\langle r\rangle \le \beta_G$, define

$$\Psi_G(r) := 1 + \big\lfloor C_G\big(R_G - r\big)\big\rfloor \in \mathbb{Z}_{\ge 1}. \tag{36}$$

Lemma E.4 (Order-preserving integerization)**.** *Fix $G \in \mathcal{G}_{\mathrm{PI}}$. If $r, r' \in \mathbb{Q}$ satisfy*

$$\langle r\rangle \le \beta_G, \qquad \langle r'\rangle \le \beta_G, \qquad r' < r,$$

then

$$\Psi_G(r') > \Psi_G(r).$$

Moreover, for every r with $\langle r\rangle \le \beta_G$, the integer $\Psi_G(r)$ has encoding length polynomial in n_G.

Proof. By Lemma E.2,

$$r - r' > 2^{-2\beta_G}.$$

Hence

$$C_G\big((R_G - r') - (R_G - r)\big) = C_G(r - r') > 2^{2\beta_G+1} \cdot 2^{-2\beta_G} = 2.$$

Therefore

$$\lfloor C_G(R_G - r') \rfloor > \lfloor C_G(R_G - r) \rfloor,$$

which implies $\Psi_G(r') > \Psi_G(r)$.

For the size bound, note that $\langle r \rangle \le \beta_G$ implies $|r| < 2^{\beta_G}$. Therefore

$$R_G - r < 2^{\beta_G+1} + 2^{\beta_G} = 3 \cdot 2^{\beta_G},$$

and similarly $R_G - r > 2^{\beta_G}$. Thus

$$1 \le \Psi_G(r) \le 1 + 3 \cdot 2^{3\beta_G+1},$$

so the binary encoding length of $\Psi_G(r)$ is $O(\beta_G)$, hence polynomial in n_G. □

E.4. The Local-Search Instance Associated with a Game

We now fix a strict Bellman-Lyapunov descent pair (F, L) relative to the bit budget b, in the sense of Definition 2.7.

Definition E.5 (Local-search specification associated with G)**.** For a game $G \in \mathcal{G}_{\mathrm{PI}}$, define the following data:

1. the finite search space
$$\mathcal{X}_b(G) = \{0,1\}^{\ell_b(G)};$$
2. the canonical initial point
$$I_G := x^{\circ}_{G,b} \in \mathcal{V}_b(G),$$
from Definition D.10;
3. the successor map
$$S_G : \mathcal{X}_b(G) \to \mathcal{X}_b(G),$$
defined by
$$S_G(x) := \begin{cases} x^{\circ}_{G,b}, & \text{if } x \notin \mathcal{V}_b(G), \\ \mathrm{enc}_{G,b}\big(F_G(\pi)\big), & \text{if } x \in \mathcal{V}_b(G) \text{ and } \pi = \mathrm{dec}_{G,b}(x); \end{cases}$$
4. the integer objective function
$$\Phi_G : \mathcal{X}_b(G) \to \mathbb{Z}_{\ge 0},$$
defined by
$$\Phi_G(x) := \begin{cases} 0, & \text{if } x \notin \mathcal{V}_b(G), \\ \Psi_G\big(L_G(\pi)\big), & \text{if } x \in \mathcal{V}_b(G) \text{ and } \pi = \mathrm{dec}_{G,b}(x), \end{cases}$$
where Ψ_G is the integerization map from Definition E.3.

Lemma E.6 (Polynomial-time computability of the local-search specification)**.** *For every $G \in \mathcal{G}_{\mathrm{PI}}$, the objects*

$$I_G, \qquad S_G, \qquad \Phi_G$$

from Definition E.5 are computable in time polynomial in $\langle G \rangle$. Moreover, every value $\Phi_G(x)$ has encoding length polynomial in $\langle G \rangle$.

Proof. The point $I_G = x^\circ_{G,b}$ is computable in polynomial time by Lemma D.11.

To compute $S_G(x)$ or $\Phi_G(x)$, first test whether x is valid. By Lemma D.8, validity testing and decoding are polynomial-time tasks in $\langle G\rangle + b_G$, hence polynomial in $\langle G\rangle$ because b is polynomial.

If x is invalid, then $S_G(x) = x^\circ_{G,b}$ and $\Phi_G(x) = 0$, both trivially computable.

Suppose now that $x \in \mathcal{V}_b(G)$, and let $\pi = \mathrm{dec}_{G,b}(x) \in \Pi_b(G)$. By Lemma 2.8, the Bellman jet $J_G(\pi)$ is exactly computable in polynomial time and has polynomial encoding length. Since F is Bellman-local, the successor encoding

$$\mathrm{enc}_{G,b}(F_G(\pi))$$

is computable in polynomial time from the encoded Bellman jet, so $S_G(x)$ is polynomial-time computable.

Similarly, since L is Bellman-local, the rational witness value $L_G(\pi)$ is polynomial-time computable from the Bellman jet. By Lemma E.1, its encoding length is at most β_G. Hence we can compute $\Psi_G(L_G(\pi))$ exactly in polynomial time by exact integer arithmetic, since R_G and C_G have encoding lengths $O(\beta_G)$, and the floor of a rational number is obtained by exact integer division. The polynomial output bound for $\Phi_G(x)$ follows from Lemma E.4. □

E.5. Characterization of Local Optima

We now prove that local optima of the above successor-objective pair are *exactly* bounded-bit stationary Nash equilibria.

Lemma E.7 (Invalid encodings are never local optima)**.** *Let $x \in \mathcal{X}_b(G) \setminus \mathcal{V}_b(G)$ be invalid. Then*

$$\Phi_G(S_G(x)) > \Phi_G(x).$$

In particular, no invalid string is a local optimum.

Proof. By Definition E.5,

$$S_G(x) = x^\circ_{G,b}, \qquad \Phi_G(x) = 0.$$

Since $x^\circ_{G,b} \in \mathcal{V}_b(G)$ by Lemma D.11, we have

$$\Phi_G(x^\circ_{G,b}) = \Psi_G\big(L_G(\pi^\circ)\big),$$

where $\pi^\circ \in \Pi_b(G)$. By Lemma E.1, $\langle L_G(\pi^\circ)\rangle \le \beta_G$, so Definition E.3 applies and yields

$$\Phi_G(x^\circ_{G,b}) = \Psi_G\big(L_G(\pi^\circ)\big) \ge 1.$$

Hence

$$\Phi_G(S_G(x)) = \Phi_G(x^\circ_{G,b}) \ge 1 > 0 = \Phi_G(x).$$

□

Lemma E.8 (Strict Bellman descent becomes strict local improvement)**.** *Let $x \in \mathcal{V}_b(G)$ be a valid encoding and $\pi = \mathrm{dec}_{G,b}(x) \in \Pi_b(G)$. If $\pi \notin \mathrm{SNE}_b(G)$, then*

$$\Phi_G(S_G(x)) > \Phi_G(x).$$

Proof. Because (F, L) is a strict Bellman-Lyapunov descent pair and $\pi \notin \mathrm{SNE}_b(G)$, we have

$$L_G(F_G(\pi)) < L_G(\pi).$$

Both rationals $L_G(F_G(\pi))$ and $L_G(\pi)$ have encoding length at most β_G by Lemma E.1. Therefore Lemma E.4 implies

$$\Psi_G\big(L_G(F_G(\pi))\big) > \Psi_G\big(L_G(\pi)\big).$$

Since x is valid, Definition E.5 gives

$$S_G(x) = \mathrm{enc}_{G,b}(F_G(\pi)),$$

and this string is valid because $F_G(\pi) \in \Pi_b(G)$. Therefore

$$\Phi_G(S_G(x)) = \Psi_G\big(L_G(F_G(\pi))\big) > \Psi_G\big(L_G(\pi)\big) = \Phi_G(x).$$

□

Lemma E.9 (Every local optimum decodes to an exact stationary Nash equilibrium)**.** *Let $x \in \mathcal{X}_b(G)$ satisfy*

$$\Phi_G(S_G(x)) \le \Phi_G(x).$$

Then $x \in \mathcal{V}_b(G)$, and

$$\mathrm{dec}_{G,b}(x) \in \mathrm{SNE}_b(G).$$

Proof. By Lemma E.7, x must be valid. Let

$$\pi := \mathrm{dec}_{G,b}(x) \in \Pi_b(G).$$

If $\pi \notin \mathrm{SNE}_b(G)$, then Lemma E.8 yields

$$\Phi_G(S_G(x)) > \Phi_G(x),$$

contradicting the local-optimality condition. Hence $\pi \in \mathrm{SNE}_b(G)$. □

E.6. PLS Membership and the Complexity Collapse

We can now prove the main barrier theorem in its bounded-bit form.

Theorem E.10 (Bellman-local strict descent implies PLS membership)**.** *Fix a polynomial bit-budget function b, and suppose (F, L) is a strict Bellman-Lyapunov descent pair relative to b on $\mathcal{G}_{\mathrm{PI}}$. Then the search problem*

$$\mathsf{GridSNE}^{b}_{\mathrm{PI}}: \quad G \mapsto \textit{find any } \pi \in \mathrm{SNE}_b(G)$$

belongs to PLS.

Proof. By Proposition D.13, $\mathrm{SNE}_b(G) \neq \varnothing$ for every G, because a strict descent pair cannot exist on an empty target set.

For an input game G, consider the local-search specification

$$\big(\mathcal{X}_b(G),\ I_G,\ S_G,\ \Phi_G\big)$$

from Definition E.5. By Lemma E.6, the initial point I_G, the successor map S_G, and the integer objective Φ_G are all computable in polynomial time in $\langle G \rangle$, and all objective values have polynomial encoding length. Thus this specification is a valid PLS instance in the standard sense of polynomial local search (Johnson et al., 1988).

A solution to this PLS instance is any string $x \in \mathcal{X}_b(G)$ satisfying

$$\Phi_G(S_G(x)) \le \Phi_G(x).$$

By Lemma E.9, every such local optimum is valid and decodes to a policy

$$\pi = \mathrm{dec}_{G,b}(x) \in \mathrm{SNE}_b(G).$$

Therefore the map

$$G \longmapsto \big(\mathcal{X}_b(G),\ I_G,\ S_G,\ \Phi_G\big)$$

is a polynomial-time reduction from $\mathsf{GridSNE}^{b}_{\mathrm{PI}}$ to the canonical PLS local-optimum problem. Hence

$$\mathsf{GridSNE}^{b}_{\mathrm{PI}} \in \mathrm{PLS}.$$

□

Corollary E.11 (Exact stationary Nash belongs to PLS)**.** *Under the hypothesis of Theorem E.10, the exact stationary Nash search problem*

$$\mathsf{StatNE}_{\mathrm{PI}}: \quad G \mapsto \textit{find any } \pi \in \mathrm{SNE}(G)$$

belongs to PLS.

Proof. By Theorem E.10, there is a PLS procedure that outputs some

$$\pi \in \mathrm{SNE}_b(G).$$

Since $\mathrm{SNE}_b(G) \subseteq \mathrm{SNE}(G)$ by definition, the same procedure solves the exact stationary Nash search problem. □

Corollary E.12 (Complexity collapse)**.** *Under the hypothesis of Theorem E.10,*

$$\mathrm{PPAD} = \mathrm{CLS}.$$

Proof. By Corollary E.11, the exact stationary Nash search problem $\mathsf{StatNE}_{\mathrm{PI}}$ lies in PLS. On the other hand, Hansen and Nie prove that computing an exact stationary Nash equilibrium in two-player discounted perfect-information stochastic games is in PPAD, which together with the earlier PPAD-hardness result of Daskalakis, Golowich, and Zhang yields PPAD-completeness for this problem (Hansen & Nie, 2025; Daskalakis et al., 2023). Since $\mathsf{StatNE}_{\mathrm{PI}}$ is PPAD-complete and also lies in PLS, it follows that

$$\mathrm{PPAD} \subseteq \mathrm{PLS}.$$

Therefore

$$\mathrm{PPAD} = \mathrm{PPAD} \cap \mathrm{PLS}.$$

Finally, Fearnley, Goldberg, Hollender, and Savani prove that

$$\mathrm{CLS} = \mathrm{PPAD} \cap \mathrm{PLS}$$

(Fearnley et al., 2022). Hence

$$\mathrm{PPAD} = \mathrm{CLS}.$$

□

Remark E.13. The reduction above is genuinely Bellman-local rather than merely stage-local. By Lemma 2.8, the hypothetical update rule is allowed to inspect the *exact* value vectors, exact one-step continuation values, exact deviation gains, and exact discounted occupancy weights of the current stationary profile. Thus the barrier persists even under exact policy evaluation and exact local exploitability information.

F. Proof of the Quantitative Strengthening

This appendix proves a quantitative strengthening of the main barrier theorem. The qualitative result in Theorem 3.1 says that universal Bellman-local strict descent would place exact stationary Nash in PLS. Here we show that if the same descent admits a *polynomially bounded Lyapunov range* together with an *inverse-polynomial one-step improvement guarantee*, then simple iteration of the Bellman-local update map yields a deterministic polynomial-time algorithm. We formulate the argument abstractly relative to a target family of stationary profiles, and then instantiate it for exact and approximate stationary Nash equilibrium.

F.1. Quantitative Bellman-Lyapunov Descent

We use FP to denote the class of total search problems solvable by a deterministic polynomial-time algorithm.

Definition F.1 (Polynomial-time solvable search problem)**.** A total search problem $(\mathcal{I}, R)$ belongs to FP if there exists a deterministic algorithm running in time polynomial in the input length such that, on every input $x \in \mathcal{I}$, the algorithm outputs some $y \in \mathrm{Sol}_R(x)$.

Fix a polynomial bit-budget function b, and for an input game G write

$$b_G := b(\langle G \rangle), \qquad n_G := \langle G \rangle + b_G.$$

Definition F.2 (Quantitative Bellman-Lyapunov descent pair)**.** Let $\mathcal{T} = \{\mathcal{T}(G)\}_{G \in \mathcal{G}_{\mathrm{PI}}}$ be a family of target sets with

$$\mathcal{T}(G) \subseteq \Pi_b(G) \qquad \text{for every } G \in \mathcal{G}_{\mathrm{PI}}.$$

A pair (F, L) is called a *quantitative Bellman-Lyapunov descent pair relative to* $\mathcal{T}$ if:

1. $F = \{F_G\}$ is a Bellman-local update map and $L = \{L_G\}$ is a Bellman-local Lyapunov witness, both relative to b;
2. there exist polynomials $B, q : \mathbb{N} \to \mathbb{N}$ such that for every $G \in \mathcal{G}_{\mathrm{PI}}$ and every $\pi \in \Pi_b(G)$,
$$0 \le L_G(\pi) \le B(n_G); \tag{37}$$
3. for every $G \in \mathcal{G}_{\mathrm{PI}}$ and every $\pi \in \mathcal{T}(G)$,
$$F_G(\pi) = \pi; \tag{38}$$
4. for every $G \in \mathcal{G}_{\mathrm{PI}}$ and every $\pi \in \Pi_b(G) \setminus \mathcal{T}(G)$,
$$L_G(\pi) - L_G(F_G(\pi)) \ge \frac{1}{q(n_G)}. \tag{39}$$

Remark F.3. A quantitative Bellman-Lyapunov descent pair is strictly stronger than the qualitative strict descent pair used in Theorem 3.1. In the latter, descent is merely strict; here it is uniformly lower bounded by an inverse-polynomial amount, while the witness itself is confined to a polynomially bounded interval.

F.2. A Generic Polynomial Hitting-Time Lemma

Proposition F.4 (Polynomial hitting time)**.** *Let (F, L) be a quantitative Bellman-Lyapunov descent pair relative to $\mathcal{T}$. Fix $G \in \mathcal{G}_{\mathrm{PI}}$, define the orbit*
$$\pi^{(t+1)} := F_G(\pi^{(t)}), \qquad t \ge 0,$$
from an arbitrary initial point $\pi^{(0)} \in \Pi_b(G)$, and set
$$T_G := B(n_G)\, q(n_G) + 1. \tag{40}$$
Then
$$\pi^{(T_G)} \in \mathcal{T}(G).$$
In particular, the orbit enters $\mathcal{T}(G)$ within at most T_G steps and, once it enters, remains there forever.

Proof. Suppose for contradiction that
$$\pi^{(t)} \notin \mathcal{T}(G) \qquad \text{for all } t = 0, 1, \dots, T_G.$$
Then by (39),
$$L_G(\pi^{(t)}) - L_G(\pi^{(t+1)}) \ge \frac{1}{q(n_G)} \qquad \text{for all } t = 0, 1, \dots, T_G - 1.$$
Summing these inequalities telescopically yields
$$L_G(\pi^{(0)}) - L_G(\pi^{(T_G)}) \ge \frac{T_G}{q(n_G)}.$$
By the definition of T_G,
$$\frac{T_G}{q(n_G)} > B(n_G).$$
On the other hand, the range bound (37) implies
$$0 \le L_G(\pi^{(T_G)}) \le L_G(\pi^{(0)}) \le B(n_G),$$
hence
$$L_G(\pi^{(0)}) - L_G(\pi^{(T_G)}) \le B(n_G),$$
a contradiction. Therefore $\pi^{(T_G)} \in \mathcal{T}(G)$.

Finally, if $\pi^{(t)} \in \mathcal{T}(G)$ for some t, then by (38),
$$\pi^{(t+1)} = F_G(\pi^{(t)}) = \pi^{(t)},$$
so the orbit remains constant thereafter. □

Proposition F.5 (Polynomial-time orbit simulation)**.** *Let (F, L) be a quantitative Bellman-Lyapunov descent pair relative to $\mathcal{T}$. Then there exists a deterministic algorithm that, on input $G \in \mathcal{G}_{\mathrm{PI}}$, computes $\pi^{(T_G)}$ in time polynomial in $\langle G \rangle$, where $\pi^{(0)} = \pi^\circ$ is the dummy policy from Definition D.10 and T_G is given by (40).*

Proof. By Lemma D.11, the dummy policy $\pi^\circ \in \Pi_b(G)$ is polynomial-time computable from G. Fix an iterate $\pi^{(t)} \in \Pi_b(G)$. By Lemma 2.8, the Bellman jet $J_G(\pi^{(t)})$ is exactly computable in time polynomial in n_G, and every coordinate of the jet has encoding length polynomial in n_G. Since F is Bellman-local, $\pi^{(t+1)} = F_G(\pi^{(t)})$ is then computable in time polynomial in n_G. Because F_G maps $\Pi_b(G)$ to itself, all iterates remain in the bounded-bit grid.

The number of iterations is

$$T_G = B(n_G)\, q(n_G) + 1,$$

which is polynomial in n_G, hence polynomial in $\langle G \rangle$ because b is polynomial. Therefore repeated exact simulation of the orbit for T_G steps runs in deterministic polynomial time. □

F.3. Exact Equilibrium Consequence

We first instantiate the generic target family to exact stationary Nash on the bounded-bit grid:

$$\mathcal{T}_{\mathrm{exact}}(G) := \mathrm{SNE}_b(G).$$

Theorem F.6 (Quantitative exact descent implies deterministic polynomial time)**.** *Fix a polynomial bit-budget function b, and suppose there exists a quantitative Bellman-Lyapunov descent pair relative to $\mathcal{T}_{\mathrm{exact}}$. Then*

$$\mathsf{GridSNE}^b_{\mathrm{PI}} \in \mathrm{FP}.$$

Consequently,

$$\mathsf{StatNE}_{\mathrm{PI}} \in \mathrm{FP}.$$

Proof. Given an input G, compute the dummy policy $\pi^\circ \in \Pi_b(G)$, and then compute the iterate $\pi^{(T_G)}$ of the Bellman-local update map F_G after $T_G = B(n_G)q(n_G) + 1$ steps. By Proposition F.5, this entire procedure runs in deterministic polynomial time.

By Proposition F.4, the output satisfies

$$\pi^{(T_G)} \in \mathcal{T}_{\mathrm{exact}}(G) = \mathrm{SNE}_b(G).$$

Hence this algorithm solves $\mathsf{GridSNE}^b_{\mathrm{PI}}$, proving $\mathsf{GridSNE}^b_{\mathrm{PI}} \in \mathrm{FP}$.

Since every output in $\mathrm{SNE}_b(G)$ is in particular an exact stationary Nash equilibrium, the same algorithm also solves the unrestricted exact stationary-Nash search problem $\mathsf{StatNE}_{\mathrm{PI}}$. □

Corollary F.7 (PPAD collapses to deterministic polynomial time)**.** *Under the hypothesis of Theorem F.6,*

$$\mathrm{PPAD} \subseteq \mathrm{FP}.$$

Equivalently, every problem in PPAD *is solvable in deterministic polynomial time.*

Proof. By Theorem F.6, $\mathsf{StatNE}_{\mathrm{PI}} \in \mathrm{FP}$. By Theorem A.11, $\mathsf{StatNE}_{\mathrm{PI}}$ is PPAD-complete. Therefore every problem in PPAD polynomial-time reduces to a search problem in FP, and is thus itself solvable in deterministic polynomial time. □

F.4. Constant-Accuracy Consequence

The previous theorem already yields an exact polynomial-time algorithm under a quantitative Lyapunov hypothesis. We now record a slightly more flexible approximate variant, which interfaces directly with the explicit constant-accuracy hardness threshold of Hansen & Nie (2025).

For $\varepsilon \geq 0$, define the bounded-bit approximate-equilibrium target set

$$\mathrm{SNE}^\varepsilon_b(G) := \{\pi \in \Pi_b(G) : \pi \text{ is a statewise stationary } \varepsilon\text{-Nash equilibrium}\}.$$

Definition F.8 (Bounded-bit approximate stationary Nash search problem)**.** Fix a polynomial bit-budget function b and $\varepsilon \geq 0$. The search problem

$$\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon}$$

takes as input a game $G \in \mathcal{G}_{\mathrm{PI}}$ and asks for any $\pi \in \mathrm{SNE}_b^{\varepsilon}(G)$.

Theorem F.9 (Quantitative ε-descent implies deterministic polynomial time)**.** *Fix a polynomial bit-budget function b and a constant $\varepsilon \geq 0$. Suppose there exists a quantitative Bellman-Lyapunov descent pair relative to the target family*

$$\mathcal{T}_{\varepsilon}(G) := \mathrm{SNE}_b^{\varepsilon}(G).$$

Then

$$\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon} \in \mathrm{FP}.$$

Proof. Apply Proposition F.5 to compute the iterate $\pi^{(T_G)}$ after $T_G = B(n_G)q(n_G) + 1$ Bellman-local updates from the dummy policy π°. By Proposition F.4,

$$\pi^{(T_G)} \in \mathcal{T}_{\varepsilon}(G) = \mathrm{SNE}_b^{\varepsilon}(G),$$

and the total running time is polynomial in $\langle G \rangle$. Hence $\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon} \in \mathrm{FP}$. □

Corollary F.10 (Explicit constant-accuracy collapse)**.** *Let*

$$\bar{\varepsilon} := \frac{3 - 2\sqrt{2}}{288}.$$

Fix any constant ε with $0 \leq \varepsilon < \bar{\varepsilon}$. If there exists a quantitative Bellman-Lyapunov descent pair relative to $\mathcal{T}_{\varepsilon}$, then

$$\mathrm{PPAD} \subseteq \mathrm{FP}.$$

Proof. By Theorem F.9, $\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon} \in \mathrm{FP}$. Hansen and Nie prove that for every

$$0 \leq \varepsilon < \frac{3 - 2\sqrt{2}}{288},$$

computing an ε-approximate Nash equilibrium in two-player $1/2$-discounted perfect-information stochastic games is PPAD-hard, even when players alternate control and each player has at most two actions per state (Hansen & Nie, 2025, Theorem 2). Therefore a deterministic polynomial-time algorithm for $\mathsf{GridSNE}_{\mathrm{PI}}^{b,\varepsilon}$ would imply that every problem in PPAD is solvable in deterministic polynomial time. □

Remark F.11. The quantitative strengthening is conceptually sharper than the PLS-barrier of Theorem 3.1. The latter rules out a universal Bellman-local strict-descent interpretation unless $\mathrm{PPAD} = \mathrm{CLS}$; the present appendix rules out the stronger possibility that such a descent dynamic enjoys a polynomial witness range and uniform inverse-polynomial progress. Under that stronger hypothesis, mere iteration of the Bellman-local update rule would solve a PPAD-hard equilibrium problem in deterministic polynomial time.

G. Bellman-Sensitivity and Genuine Dynamicity

This appendix formalizes the sense in which the Bellman-local model used in the main theorem is genuinely dynamic. The key point is that Bellman-local update rules may depend on continuation values and discounted occupancies, and therefore can react to downstream transition-reward structure even when the entire *local stage data* at the current state are held fixed. We make this precise by defining a weak notion of stage-locality at a state and then exhibiting a Bellman-local update family that is not stage-local.

G.1. Stage Signatures and Stage-Local Update Rules

In this appendix, when comparing two games, we only compare games defined on the same labeled state space S, with the same label $s \in S$ singled out. This is sufficient for our purpose: we want to show that even when the current state and its outgoing stage data are literally identical, a Bellman-local update may still change because of downstream continuation structure.

Definition G.1 (Stage signature at a state)**.** Let $G, \widetilde{G} \in \mathcal{G}_{\mathrm{PI}}$ be games on the same labeled state space S, and let $s \in S$. Let $\pi \in \Pi(G)$ and $\widetilde{\pi} \in \Pi(\widetilde{G})$ be stationary profiles. We say that (G, π) and $(\widetilde{G}, \widetilde{\pi})$ have the *same stage signature at* s if all of the following hold:

1. the controller of s is the same in both games:
$$c_G(s) = c_{\widetilde{G}}(s);$$
2. the action set at s is the same in both games:
$$A_G(s) = A_{\widetilde{G}}(s);$$
3. the local mixed action at s is the same:
$$\pi(\cdot \mid s) = \widetilde{\pi}(\cdot \mid s);$$
4. the immediate stage rewards at s agree coordinatewise: for every player $i \in \{1, 2\}$ and every $a \in A_G(s)$,
$$r_i^G(s, a) = r_i^{\widetilde{G}}(s, a);$$
5. the one-step transition law out of s agrees coordinatewise: for every $a \in A_G(s)$,
$$P^G(\cdot \mid s, a) = P^{\widetilde{G}}(\cdot \mid s, a) \quad \text{as distributions on } S.$$

Remark G.2. The stage signature records precisely the data visible at the current state in a one-shot stage-game view: who acts, which actions are available, which local mixed action is currently played, what the immediate rewards are, and where one moves in one step. By design, it does *not* include any information about the rewards or transitions *at successor states*.

Definition G.3 (Stage-locality at a state)**.** A family of stationary-policy update maps
$$U = \{U_G : \Pi(G) \to \Pi(G)\}_{G \in \mathcal{G}_{\mathrm{PI}}}$$
is *stage-local at a state* s if for every pair of games $G, \widetilde{G} \in \mathcal{G}_{\mathrm{PI}}$ on the same labeled state space S, and every pair of stationary profiles $\pi \in \Pi(G)$, $\widetilde{\pi} \in \Pi(\widetilde{G})$, the implication
$$(G, \pi) \text{ and } (\widetilde{G}, \widetilde{\pi}) \text{ have the same stage signature at } s$$
implies
$$U_G(\pi)(\cdot \mid s) = U_{\widetilde{G}}(\widetilde{\pi})(\cdot \mid s).$$

Stage-locality is intentionally weak: it only requires that the update at s be determined by the local stage data at s. Any update rule whose decision at s depends on downstream continuation values or downstream transition-reward structure fails this property.

G.2. A Bellman-Local Update Family that is not Stage-Local

We now construct an explicit Bellman-local update family whose behavior at the current state depends on downstream continuation structure even when the stage signature is fixed.

Fix once and for all the input order on states and actions used in Section D. For any game $G \in \mathcal{G}_{\mathrm{PI}}$, let $s^\star(G)$ be the first state in the input order satisfying:
$$s^\star(G) \in S_1 \qquad \text{and} \qquad |A(s^\star(G))| \geq 2,$$
if such a state exists. If no such state exists, we leave the policy unchanged.

When $s^\star(G)$ exists, let
$$a_1^\star(G),\, a_2^\star(G) \in A(s^\star(G))$$
be the first two actions in the input order at $s^\star(G)$.

Definition G.4 (A Bellman-greedy update family)**.** Define the family $F^{\mathrm{BG}} = \{F_G^{\mathrm{BG}}\}_{G \in \mathcal{G}_{\mathrm{PI}}}$ as follows. For a stationary profile $\pi \in \Pi(G)$:

1. if G has no player-1 state with at least two actions, set
$$F_G^{\mathrm{BG}}(\pi) := \pi;$$
2. otherwise, let $s^\star = s^\star(G)$ and $a_1^\star, a_2^\star$ be as above, and define $F_G^{\mathrm{BG}}(\pi)$ by keeping all policy coordinates unchanged except at $s^\star$, where:
$$F_G^{\mathrm{BG}}(\pi)(\cdot \mid s^\star) := \begin{cases} \delta_{a_1^\star}, & \text{if } Q_1^\pi(s^\star, a_1^\star) \geq Q_1^\pi(s^\star, a_2^\star), \\ \delta_{a_2^\star}, & \text{if } Q_1^\pi(s^\star, a_1^\star) < Q_1^\pi(s^\star, a_2^\star), \end{cases}$$
where δ_a denotes the pure distribution concentrated on a.

Lemma G.5. *The family F^{BG} is Bellman-local in the sense of Definition 2.5. Moreover, if $\pi \in \Pi_b(G)$ for any bit budget b, then*
$$F_G^{\mathrm{BG}}(\pi) \in \Pi_b(G).$$

Proof. To compute $F_G^{\mathrm{BG}}(\pi)$, it suffices to read from the game input the first player-1 state $s^\star$ with at least two actions and the first two actions $a_1^\star, a_2^\star$ at that state, then compare the two exact Bellman jet coordinates
$$Q_1^\pi(s^\star, a_1^\star) \qquad \text{and} \qquad Q_1^\pi(s^\star, a_2^\star).$$
This is polynomial-time computable from the Bellman jet $J_G(\pi)$, so F^{BG} is Bellman-local.

If $\pi \in \Pi_b(G)$, then $F_G^{\mathrm{BG}}(\pi)$ changes at most one state, and at that state replaces the mixed action by a pure distribution. Since the probabilities 0 and 1 belong to every bounded-bit grid $\Pi_b(G)$ by Remark D.2, the updated policy remains in $\Pi_b(G)$. □

We now exhibit two games on which F^{BG} reacts differently at the same current state even though the local stage signature at that state is identical.

Proposition G.6 (Bellman sensitivity to downstream structure)**.** *There exist two games $G^+, G^- \in \mathcal{G}_{\mathrm{PI}}$ on the same labeled state space $S = \{s, t_a, t_b\}$, and a common stationary policy profile $\pi \in \Pi(G^+) \cap \Pi(G^-)$, such that:*

1. *(G^+, π) and (G^-, π) have the same stage signature at s;*
2. *the Bellman-greedy family F^{BG} satisfies*
$$F_{G^+}^{\mathrm{BG}}(\pi)(\cdot \mid s) \neq F_{G^-}^{\mathrm{BG}}(\pi)(\cdot \mid s).$$

Consequently, F^{BG} is not stage-local at s.

Proof. We define the two games G^+ and G^- on the common state space
$$S = \{s, t_a, t_b\},$$
with controller partition
$$S_1 = \{s\}, \qquad S_2 = \{t_a, t_b\}.$$
The action sets are
$$A(s) = \{a, b\}, \qquad A(t_a) = A(t_b) = \{*\}.$$
At the distinguished state s, both games have identical one-step dynamics:
$$P(\, t_a \mid s, a\,) = 1, \qquad P(\, t_b \mid s, b\,) = 1,$$

and no other successor states occur. The unique action $*$ at t_a and t_b is absorbing in both games:

$$P(\, t_a \mid t_a, * \,) = 1, \qquad P(\, t_b \mid t_b, * \,) = 1.$$

The immediate rewards at s are identical and equal to zero for both players:

$$r_i^{G^+}(s,a) = r_i^{G^+}(s,b) = r_i^{G^-}(s,a) = r_i^{G^-}(s,b) = 0, \qquad i \in \{1,2\}.$$

Player 2's rewards are irrelevant; set them to zero everywhere in both games. The only difference between G^+ and G^- is player 1's reward at the absorbing successor states:

$$r_1^{G^+}(t_a, *) = 1, \qquad r_1^{G^+}(t_b, *) = 0,$$

while

$$r_1^{G^-}(t_a, *) = 0, \qquad r_1^{G^-}(t_b, *) = 1.$$

Define the common stationary policy π by

$$\pi(a \mid s) = \pi(b \mid s) = \frac{1}{2}, \qquad \pi(* \mid t_a) = \pi(* \mid t_b) = 1.$$

We first verify item 1. At the state s, both games have:

- the same controller $c(s) = 1$,
- the same action set $A(s) = \{a, b\}$,
- the same local mixed action $\pi(\cdot \mid s) = (1/2, 1/2)$,
- the same immediate rewards $r_i(s,a) = r_i(s,b) = 0$ for both players,
- the same outgoing transition laws $P(\cdot \mid s, a) = \delta_{t_a}$ and $P(\cdot \mid s, b) = \delta_{t_b}$.

Thus (G^+, π) and (G^-, π) have the same stage signature at s.

We now compute the Bellman continuation values for player 1. In G^+, the state t_a is absorbing with perpetual stage reward 1, hence by the normalized discounted payoff convention,

$$V_1^\pi(t_a) = (1-\gamma) \sum_{k=0}^{\infty} \gamma^k \cdot 1 = 1.$$

Similarly,

$$V_1^\pi(t_b) = (1-\gamma) \sum_{k=0}^{\infty} \gamma^k \cdot 0 = 0.$$

Since the immediate reward at s is 0,

$$Q_1^\pi(s,a) = (1-\gamma)\cdot 0 + \gamma V_1^\pi(t_a) = \gamma, \qquad Q_1^\pi(s,b) = (1-\gamma)\cdot 0 + \gamma V_1^\pi(t_b) = 0.$$

Therefore

$$Q_1^\pi(s,a) > Q_1^\pi(s,b)$$

in G^+, so by Definition G.4,

$$F_{G^+}^{\mathrm{BG}}(\pi)(\cdot \mid s) = \delta_a.$$

In G^-, the roles of t_a and t_b are reversed, so

$$V_1^\pi(t_a) = 0, \qquad V_1^\pi(t_b) = 1,$$

and consequently

$$Q_1^\pi(s,a) = 0, \qquad Q_1^\pi(s,b) = \gamma.$$

Hence

$$Q_1^\pi(s,b) > Q_1^\pi(s,a),$$

so

$$F^{\mathrm{BG}}_{G^-}(\pi)(\cdot \mid s) = \delta_b.$$

Thus

$$F^{\mathrm{BG}}_{G^+}(\pi)(\cdot \mid s) = \delta_a \neq \delta_b = F^{\mathrm{BG}}_{G^-}(\pi)(\cdot \mid s),$$

proving item 2. By Definition G.3, the family F^{BG} is not stage-local at s. □

Corollary G.7 (Bellman-locality strictly exceeds stage-locality)**.** *There exists a Bellman-local update family that cannot be represented as a stage-local update family at every state.*

Proof. The family F^{BG} is Bellman-local by Lemma G.5, but it is not stage-local at the state s by Proposition G.6. Therefore it cannot be stage-local at every state. □

Remark G.8. Proposition G.6 isolates the precise sense in which the barrier theorem is dynamic. The two games G^+ and G^- are indistinguishable from the viewpoint of the local stage data at s, yet the Bellman-local update moves in opposite directions because the continuation value of the two successor states has been reversed. Thus Bellman-locality is not a cosmetic reformulation of a one-shot local rule; it genuinely exploits the stochastic dynamic structure.

H. Compatibility with Positive Structured Subclasses

This appendix explains why the barrier theorems proved in Sections 3 and 4 do not contradict the positive algorithmic and convergence results known for structured subclasses of Markov games. The key logical point is simple but important: our barrier is *universal*. It rules out Bellman-local descent principles defined on the entire class $\mathcal{G}_{\mathrm{PI}}$ and targeting *exact* stationary Nash equilibrium. Once one restricts the domain, relaxes the target, or weakens the certificate, the conclusion changes accordingly.

H.1. Restricted Subclasses and Restricted Search Problems

Definition H.1 (Structured subclass)**.** A *structured subclass* is any set

$$\mathfrak{C} \subseteq \mathcal{G}_{\mathrm{PI}}$$

of discounted perfect-information stochastic games.

Given a structured subclass $\mathfrak{C}$, define the restricted exact search problem

$$\mathsf{StatNE}_{\mathfrak{C}}: \quad \text{input } G \in \mathfrak{C}, \qquad \text{output any } \pi \in \mathrm{SNE}(G).$$

Fix also a polynomial bit-budget function b, and define

$$\mathsf{GridSNE}^{b}_{\mathfrak{C}}: \quad \text{input } G \in \mathfrak{C}, \qquad \text{output any } \pi \in \mathrm{SNE}_b(G).$$

For $\varepsilon \geq 0$, define similarly

$$\mathsf{GridSNE}^{b,\varepsilon}_{\mathfrak{C}}: \quad \text{input } G \in \mathfrak{C}, \qquad \text{output any } \pi \in \mathrm{SNE}^{\varepsilon}_b(G).$$

Proposition H.2 (Restriction principle: qualitative form)**.** *Fix a polynomial bit-budget function b, let $\mathfrak{C} \subseteq \mathcal{G}_{\mathrm{PI}}$, and suppose there exists a strict Bellman-Lyapunov descent pair (F, L) relative to b on $\mathfrak{C}$, i.e., for every $G \in \mathfrak{C}$,*

$$F_G : \Pi_b(G) \to \Pi_b(G), \qquad L_G : \Pi_b(G) \to \mathbb{Q},$$

with

$$F_G(\pi) = \pi \quad \forall \pi \in \mathrm{SNE}_b(G), \qquad L_G(F_G(\pi)) < L_G(\pi) \quad \forall \pi \in \Pi_b(G) \setminus \mathrm{SNE}_b(G).$$

Then

$$\mathsf{GridSNE}^{b}_{\mathfrak{C}} \in \mathrm{PLS}, \qquad \mathsf{StatNE}_{\mathfrak{C}} \in \mathrm{PLS}.$$

Proof. The proof is exactly the proof of Theorem E.10, but restricted to the domain $\mathfrak{C}$. For each $G \in \mathfrak{C}$, one constructs the same local-search instance on $\mathcal{X}_b(G)$ from Definition E.5. Every local optimum decodes to an element of $\mathrm{SNE}_b(G)$ by the same argument as in Lemmas E.7, E.8, and E.9. Hence $\mathsf{GridSNE}^b_{\mathfrak{C}} \in \mathrm{PLS}$. Since $\mathrm{SNE}_b(G) \subseteq \mathrm{SNE}(G)$, the same reduction solves $\mathsf{StatNE}_{\mathfrak{C}}$. □

Proposition H.3 (Restriction principle: quantitative form)**.** *Fix a polynomial bit-budget function b, a subclass $\mathfrak{C} \subseteq \mathcal{G}_{\mathrm{PI}}$, and a constant $\varepsilon \geq 0$. Suppose there exists a quantitative Bellman-Lyapunov descent pair relative to the target family*

$$\mathcal{T}_\varepsilon(G) := \mathrm{SNE}^\varepsilon_b(G), \qquad G \in \mathfrak{C},$$

with polynomial range and inverse-polynomial progress as in Definition F.2. Then

$$\mathsf{GridSNE}^{b,\varepsilon}_{\mathfrak{C}} \in \mathrm{FP}.$$

In particular, when $\varepsilon = 0$,

$$\mathsf{StatNE}_{\mathfrak{C}} \in \mathrm{FP}.$$

Proof. The proof is exactly the proof of Theorems F.6 and F.9, restricted to inputs $G \in \mathfrak{C}$. The polynomial hitting-time argument from Proposition F.4 remains valid verbatim, because it uses only the range bound, the progress inequality, and the Bellman-local computability of the update rule. □

Corollary H.4 (Why the barrier is compatible with structured positives)**.** *Theorems 3.1 and 4.1 do not conflict with any result that falls into at least one of the following categories:*

1. *the result is proved only on a strict subclass $\mathfrak{C} \subsetneq \mathcal{G}_{\mathrm{PI}}$;*
2. *the target object is not exact stationary Nash on all of $\mathcal{G}_{\mathrm{PI}}$, e.g., an approximate equilibrium, a value vector, or a best-response fixed point on a restricted domain;*
3. *the proof uses only an approximate Lyapunov surrogate, asymptotic convergence, or other guarantees weaker than a universal exact strict Bellman-Lyapunov descent pair.*

Proof. Theorem 3.1 assumes a universal strict Bellman-Lyapunov descent pair on the full class $\mathcal{G}_{\mathrm{PI}}$ and concludes a statement about the unrestricted exact search problem $\mathsf{StatNE}_{\mathrm{PI}}$. By Proposition H.2, if the same type of descent exists only on a strict subclass $\mathfrak{C}$, then the conclusion applies only to $\mathsf{StatNE}_{\mathfrak{C}}$, not to the unrestricted problem.

Similarly, Theorem 4.1 assumes a quantitative descent pair relative to a fixed target family and concludes deterministic polynomial-time solvability of the corresponding target problem. By Proposition H.3, if the target is weaker than exact stationary Nash, or the domain is restricted, then only the corresponding weaker/restricted search problem is placed in FP.

Finally, if a result does not even posit an exact strict Bellman-Lyapunov descent pair—for example because it uses only an approximate Lyapunov function, or proves asymptotic convergence under additional regularity—then the hypotheses of our main theorems are simply not met. □

H.2. Examples from the Literature

We now record the main structured positive regimes relevant to this paper.

Remark H.5 (Zero-sum and Shapley-type positive regimes). The zero-sum side of stochastic-game theory is compatible with our barrier for two independent reasons. First, it is a strict subclass of $\mathcal{G}_{\mathrm{PI}}$, so any positive result there falls under Corollary H.4. Second, the algorithmic target is often the zero-sum value or an ε-optimal strategy pair rather than the unrestricted general-sum exact stationary-Nash problem.

Concrete examples include the strongly polynomial strategy-iteration result of Hansen et al. (2013) for two-player turn-based discounted zero-sum games with constant discount factor, the near-optimal time/sample-complexity algorithm of Sidford et al. (2020) for computing ε-optimal strategies in discounted turn-based zero-sum stochastic games, and the UEOPL placement and faster approximation algorithm for Shapley games via monotone contractions obtained by Batziou et al. (2025). None of these results asserts a universal exact Bellman-local strict descent principle for all general-sum games, so none contradicts our theorem.

Remark H.6 (Equilibrium-collapse subclasses). A second family of positive results exploits structural collapse phenomena. Kalogiannis and Panageas show that in zero-sum polymatrix Markov games, coarse correlated equilibria collapse to Nash equilibria, which leads to efficient approximate-Nash computation in that subclass (Kalogiannis & Panageas, 2023). Anagnostides, Panageas, Farina, and Sandholm prove that in multi-player Markov games with a single controller, optimistic policy gradient converges to stationary ε-Nash equilibrium under the additional assumption of equilibrium collapse (Anagnostides et al., 2024).

These results are exactly of the type covered by Corollary H.4: they are proved on highly structured proper subclasses, and their target is approximate Nash under extra collapse assumptions rather than a universal exact-descent principle on all of $\mathcal{G}_{\mathrm{PI}}$.

Remark H.7 (Potential and near-potential structure). Potential structure is another major positive regime outside the scope of our barrier. Leonardos, Overman, Panageas, and Piliouras prove global convergence of independent policy gradient in Markov potential games (Leonardos et al., 2022), and Fox, McAleer, Overman, and Panageas prove last-iterate convergence of independent natural policy gradient in the same class (Fox et al., 2022). More recently, Maheshwari, Wu, and Sastry introduce a Markov near-potential function that acts as an *approximate* Lyapunov function for decentralized actor-critic dynamics in general-sum Markov games (Maheshwari et al., 2024).

Again, there is no contradiction. Markov potential games form a structured subclass, so positive convergence there is compatible with Proposition H.2. The near-potential result goes further toward general-sum games, but its certificate is deliberately weaker than the exact strict Bellman-Lyapunov descent pair ruled out by our theorem: it is an approximate Lyapunov surrogate used to describe a convergent set, not a universal exact witness for stationary Nash on all $\mathcal{G}_{\mathrm{PI}}$.

Remark H.8 (Interpretation). Appendix H should therefore be read as a frontier map. Our barrier theorem does *not* say that Bellman-local learning is impossible in dynamic games. Rather, it says that one cannot hope for a single universal Bellman-local strict-descent principle that solves exact stationary Nash across the full class $\mathcal{G}_{\mathrm{PI}}$ without causing a complexity collapse. The positive literature succeeds precisely by exploiting structure—zero-sum geometry, equilibrium collapse, potentiality, or near-potentiality—or by weakening the target from exact unrestricted stationary Nash to a structured or approximate alternative.