ON THE LIMIT EQUILIBRIUM PAYOFF SET IN REPEATED AND STOCHASTIC GAMES

By

Johannes Hörner, Satoru Takahashi and Nicolas Vieille

February 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1848

On the Limit Equilibrium Payoff Set in Repeated and Stochastic Games[*]

Johannes Hörner[†], Satoru Takahashi[‡] and Nicolas Vieille[§]

February 14, 2012

Abstract

This paper provides a dual characterization of the limit set of perfect public equilibrium payoffs in stochastic games (in particular, repeated games) as the discount factor tends to one. As a first corollary, the folk theorems of Fudenberg, Levine and Maskin (1994), Kandori and Matsushima (1998) and Hörner, Sugaya, Takahashi and Vieille (2011) obtain. As a second corollary, in the context of repeated games, it follows that this limit set of payoffs is a polytope (a bounded polyhedron) when attention is restricted to equilibria in pure strategies. We provide a two-player game in which this limit set is not a polytope when mixed strategies are considered.

Keywords: stochastic games, repeated games, folk theorem.

JEL codes: C72, C73

1 Introduction

Given how much the literature on repeated and stochastic games has focused on the case in which discounting vanishes, it might be surprising how little is known about the limiting

1

equilibrium payoff set when sufficient conditions for a folk theorem are not met. In the case of games with imperfect public monitoring, our knowledge about the limiting set of perfect public equilibrium payoffs derives from the characterizations of Fudenberg and Levine (1994), and its generalizations by Fudenberg, Levine and Takahashi (2007), and Hörner, Sugaya, Takahashi and Vieille (2011) in terms of a parameterized family of nonlinear programs: whenever the characterization applies, (i) the limit of the equilibrium payoff set is well-defined, and (ii) it is compact, convex and semialgebraic.[1]

This paper provides a characterization of this limit set that gives additional insights and results. We study the dual of the program considered in Hörner, Sugaya, Takahashi and Vieille (2011). We show that this dual program offers several advantages over the primal: (i) it admits a straightforward interpretation; (ii) because the constraint set depends on the parameters of the program through the parameters' signs only, it is easy to solve especially for repeated games; (iii) the various sufficient conditions for a folk theorem for repeated and stochastic games with public monitoring that are found in the literature obtain effortlessly.

To demonstrate the tractability of the dual program, we exploit it to establish that the limit set of pure-strategy (perfect public) equilibrium payoffs in repeated games is a polytope (whenever the characterization applies). We provide an example of a two-player game with two signals for which the limit set of perfect public equilibrium payoffs is not a polytope when mixed strategy equilibria are considered.

While our analysis focuses on the limit case in which the discount factor tends to one, duality has already been applied to the case of repeated games by Cheng (2004), who obtains a characterization for a fixed discount factor that is the counterpart of ours. To our knowledge, Cheng is the first author to use duality to characterize the set of equilibrium payoffs in repeated games. A related application of duality to incentive problems in a static context can be found in Obara and Rahman (2010). On the other hand, duality is a standard tool in Markov decision processes, the "one-player" version of a stochastic game.

2 The Dual Program

In this section, we provide a characterization of the limit payoff set in stochastic games with public signals, or more precisely, another characterization of the nonlinear programs whose

---

[1]Even less is known for Nash equilibria. For instance, convergence of the equilibrium payoff set is an open problem.

solution is key to the description of this payoff set. We follow Hörner, Sugaya, Takahashi and Vieille (2011, hereafter HSTV) for notation and assumptions. At each stage, the game is in one state, and players simultaneously choose actions. Nature then determines both the current payoff, the next state and a public signal, as a function of the current state and the action profile. The sets S of possible states, I of players, A^i of actions available to player i, and Y of public signals are assumed finite. Given an action profile $a \in A := \times_i A^i$, and a state $s \in S$, we denote by $r(s,a) \in \mathbf{R}^I$ the payoff (or reward) profile when in state s given a, and by $p(t,y|s,a)$ the joint probability of moving to state $t \in S$ and of getting the public signal $y \in Y$. A repeated game is the special case in which there is a singleton state.

At the end of each period, the only information publicly available to all players consists of nature's choices: the next state together with the public signal. When properly interpreting Y, this includes the case of perfect monitoring and the case of publicly observed payoffs.

In each period $n = 1, 2, \ldots$, the state s_n is observed, the stage game is played, the action profile a_n is realized, and the public signal y_n is then revealed. The stochastic game is parameterized by the initial state s_1. The public history at the beginning of period n is then $h_n = (s_1, y_1, \ldots, s_{n-1}, y_{n-1}, s_n)$. We set $H_1 := S$, the set of initial states. The set of public histories at the beginning of period n is therefore $H_n := (S \times Y)^{n-1} \times S$, and we let $H := \bigcup_{n \geq 1} H_n$ denote the set of all public histories. The private history for player i at the beginning of period n is a sequence $h_n^i = (s_1, a_1, y_1, \ldots, s_{n-1}, a_{n-1}, y_{n-1}, s_n)$, and we similarly define $H_1^i := S$, $H_n^i := (S \times A^i \times Y)^{n-1} \times S$ and $H^i := \bigcup_{n \geq 1} H_n^i$.

A (behavior) strategy for player $i \in I$ is a map $\sigma^i : H^i \to \Delta(A^i)$. Every pair of initial state s_1 and strategy profile σ generates a probability distribution over histories in the obvious way and thus also generates a distribution over sequences of the players' rewards. Players seek to maximize their payoff, that is, the average discounted sum of their rewards, using a common discount factor $\delta < 1$. Thus, the payoff of player $i \in I$ if the initial state is s_1 and the players follow the strategy profile σ is defined as

$$\sum_{n=1}^{\infty} (1 - \delta)\delta^{n-1} \mathbf{E}_{s_1, \sigma}[r^i(s_n, a_n)].$$

A strategy σ^i is public if it depends on the public history only, and not on player i's private information. That is, a public strategy is a map $\sigma^i : H \to \Delta(A^i)$. A *perfect public* equilibrium (hereafter, PPE, or simply equilibrium) is a profile of public strategies such that, given any

period n and public history h_n, the strategy profile is a Nash equilibrium from that period on. We denote by $E(s, \delta) \subset \mathbf{R}^I$ the (compact) set of PPE payoffs of the game with initial state $s \in S$ and discount factor $\delta < 1$. All statements about convergence of, or equality between sets are understood in the sense of the Hausdorff distance $d(A, B)$ between sets A, B.

The main element of the characterization of HSTV is the solution to the following nonlinear program, where $\lambda \in \mathbf{R}^I$ is fixed. Given a state $s \in S$ and a map $x : S \times Y \to \mathbf{R}^{S \times I}$, we denote by $\Gamma(s, x)$ the one-shot game with action sets A^i and payoff function

$$r(s, a_s) + \sum_{t \in S} \sum_{y \in Y} p(t, y | s, a_s) x_t(s, y),$$

where $x_t(s, y) \in \mathbf{R}^I$ is the t-th component of $x(s, y)$.

Given $\lambda \in \mathbf{R}^I$, we denote by $\mathcal{P}(\lambda)$ the maximization program

$$\sup_{v, x, \alpha} \lambda \cdot v,$$

where the supremum is taken over all $v \in \mathbf{R}^I$, $x : S \times Y \to \mathbf{R}^{S \times I}$, and $\alpha = (\alpha_s)_s \in (\times_{i \in I} \Delta(A^i))^S$ such that

(i) For each s, α_s is a Nash equilibrium with payoff v of the game $\Gamma(s, x)$;

(ii) For each $T \subseteq S$, for each permutation $\varphi : T \to T$ and each map $\psi : T \to Y$, one has $\lambda \cdot \sum_{s \in T} x_{\varphi(s)}(s, \psi(s)) \leq 0$.

The program $\mathcal{P}(\lambda)$ is a generalization to stochastic games of the program introduced by Fudenberg and Levine (1994) for repeated games, based in turn on the recursive representation of the payoff set given by Abreu, Pearce and Stacchetti (1990).

Denote by $k(\lambda) \in [-\infty, +\infty]$ the value of $\mathcal{P}(\lambda)$. HSTV prove that the feasible set of $\mathcal{P}(\lambda)$ is non-empty, so that $k(\lambda) > -\infty$, and that the value of $\mathcal{P}(\lambda)$ is finite, so that $k(\lambda) < +\infty$.

HSTV assume that the limit set of PPE payoffs is independent of the initial state: for all $s, t \in S$, $\lim_{\delta \to 1} d(E(s, \delta), E(t, \delta)) = 0$ (Assumption **A**). HSTV prove that, under Assumption **A** and a full-dimensionality condition, the family of programs (indexed by λ) characterizes the limit set of (perfect public) equilibrium payoffs as $\delta \to 1$,

$$\lim_{\delta \to 1} E(s, \delta) = \bigcap_{\lambda \in \mathbf{R}^I} \{v \in \mathbf{R}^I \mid \lambda \cdot v \leq k(\lambda)\} \quad \forall s \in S.$$

4

As our focus is the program itself, we shall not need Assumption **A** for what follows.

For a fixed Markov strategy $(\alpha_s)_s$, the feasible set is non-empty if and only if for all s, α_s is *admissible*, in the sense that, for all i, if there exists $\nu_s^i \in \Delta(A^i)$ such that, for all (t, y),

$$\sum_{a^i \in A^i} \nu_s^i(a^i) p(t, y | s, a^i, \alpha_s^{-i}) = p(t, y | s, \alpha_s),$$

then

$$\sum_{a^i \in A^i} \nu_s^i(a^i) r^i(s, a^i, \alpha_s^{-i}) \leq r^i(s, \alpha_s).$$

Indeed, it follows from Fan (1956) that there exists $x : S \times Y \to \mathbf{R}^{S \times I}$ such that for each s, α_s is a Nash equilibrium of the game $\Gamma(s, x)$ if and only if for each s, α_s is admissible. Adding a constant to each $x_t(s, y)$ that is independent of (t, y), we may assume that the equilibrium payoff v_s is independent of s. Finally, considering any i for which $\lambda^i \neq 0$, we may add (or subtract if $\lambda^i < 0$) a constant to $x_t^i(s, y)$, independent of s, t, y, so that the constraint **(ii)** is satisfied.

Define the program $\tilde{\mathcal{P}}(\lambda)$ as follows:

$$\sup_{\alpha \in (\times_i \Delta(A^i))^S, \alpha_s \text{ admissible}} \min \sum_{s,i} \lambda^i \beta_s r^i \left(s, \hat{\alpha}_s^i, \alpha_s^{-i} \right),$$

where the minimum is over $(\hat{\alpha}_s^i)_{s,i}$ for all i for which $\lambda^i \neq 0$, with $\hat{\alpha}_s^i \in \mathbf{R}^{A^i}$, $\sum_{a^i} \hat{\alpha}_s^i(a^i) = 1$, $\hat{\alpha}_s^i(a^i) \leq 0$ if $\lambda^i > 0$ and $\alpha_s^i(a^i) = 0$, $\hat{\alpha}_s^i(a^i) \geq 0$ if $\lambda^i < 0$ and $\alpha_s^i(a^i) = 0$, and

$$\hat{p}(t, y | s) := p \left(t, y | s, \hat{\alpha}_s^i, \alpha_s^{-i} \right) \geq 0,$$

as well as over $\beta_s \geq 0$, $\sum_s \beta_s = 1$ such that $(\beta_s)_s$ is an invariant distribution of $\hat{p}(t \times Y | s)$. (If there are multiple invariant distributions, use the one that minimizes the objective function.)

The main result of this section is the following:

Theorem 1 *For all $\lambda \in \mathbf{R}^I$, the programs $\mathcal{P}(\lambda)$ and $\tilde{\mathcal{P}}(\lambda)$ yield the same value.*

The constraints appearing in $\tilde{\mathcal{P}}(\lambda)$ have a natural interpretation: each player can only deviate to a strategy $(\hat{\alpha}_s^i)_s$ that leads to a distribution over signals and states—via the invariant distribution—that is the same for all players' deviations. That is, it is as if adversarial players were choosing the deviation strategies $(\hat{\alpha}_s^i)_{s,i}$ in a coordinated manner, subject to the constraint

5

that they cannot be told apart, whether through the public signals or through the state transitions, and the objective is to minimize the λ-weighted average payoff given those deviations. (Notice, however, that "deviations" are defined in an unusual way so that $\hat{\alpha}_s^i(a^i)$ may take negative values.)

One of the advantages of this dual characterization is that the weight vector λ no longer appears in the constraints, or rather, it only appears via the signs of these weights. This makes the program especially tractable for repeated games: for each admissible strategy profile α and each "orthant" in which λ might lie, we are left with a linear program with variables $\hat{\alpha} = (\hat{\alpha}^i)_i$, where λ appears only in the objective. Hence, for each α and each orthant, there are only finitely many candidates of $\hat{\alpha}$ to consider. This does not only make the analysis tractable, but it also yields some qualitative results. See Section 4.

Cheng (2004)'s Theorem 5 corresponds to $\tilde{\mathcal{P}}(\lambda)$ with $|S| = 1$, where $(\beta_s)_s$ collapses to the point mass.

From the dual program $\tilde{\mathcal{P}}(\lambda)$, and given the characterization of the equilibrium payoff sets from Fudenberg and Levine for the case of repeated games, the existing folk theorems follow immediately. This is obvious for the sufficient conditions given by Fudenberg, Levine and Maskin (1994). As for those of Kandori and Matsushima (1998), note that their conditions can be stated in terms of convex cones.[2] Adapting slightly their notation, let $Q^i(a) := \{p(\cdot|a^{-i}, \tilde{a}^i)|\tilde{a}^i \in A^i \setminus \{a^i\}\}$ be the set of distributions over signals as player i's action varies over all his actions but a^i. Let $C^i(a)$ denote the convex cone with vertex 0 spanned by $Q^i(a) - p(\cdot|a)$. Assumption **A2** of Kandori and Matsushima requires $C^i(a) \cap -C^j(a) = \{0\}$ and that 0 is not a non-trivial conical combination of $Q^i(a) - p(\cdot|a)$, whereas Assumption **A3** requires $C^i(a) \cap C^j(a) = \{0\}$ for all $i \neq j$ and $a \in Ex(A)$ (the set of action profiles achieving some extreme point of the feasible payoff set). Note now that the restriction on $\hat{\alpha}$, when $\alpha = a$ is pure, is that $p(\cdot|\hat{\alpha}^i, a^{-i}) - p(\cdot|a) \in -C^i(a)$ whenever $\lambda^i > 0$, and $p(\cdot|\hat{\alpha}^i, a^{-i}) - p(\cdot|a) \in C^i(a)$ whenever $\lambda^i < 0$. Assumptions **A2** and **A3** then imply that $\hat{\alpha}^i = a^i$ for non-coordinate directions λ.[3]

Similarly, HSTV's folk theorem for stochastic games follows immediately under their assumptions **F1** and **F2**. Our results are also reminiscent of the link between the average cost optimality equation and linear programming formulations in Markov decision processes (see Hernández-Lerma and Lasserre, 1999, Ch. 12), and consistent with the results of Hoffman and

---

[2]Note that the working paper of Kandori and Matsushima (1998) gives weaker conditions than the published one, which can be seen to also follow immediately from $\tilde{\mathcal{P}}(\lambda)$.

[3]For coordinate directions, admissibility suffices (cf. Kandori and Matsushima's Assumption **A1**).

Karp (1966).[4]

As a final remark, one can characterize the limit set of pure-strategy PPE payoffs by modifying the primal $\mathcal{P}(\lambda)$ so that the supremum is taken over pure strategies $\alpha \in A^S$. The corresponding dual $\tilde{\mathcal{P}}(\lambda)$ is given by taking the supremum over all admissible pure strategies.

3 Proof of Theorem 1

Fix throughout some strategy $(\alpha_s)_s$ such that α_s is admissible for all s. We can rewrite $\mathcal{P}(\lambda)$ as

$$\max_{x,v} \lambda \cdot v$$

over x and v such that, for all s, i,

$$\sum_{t,y} p\left(t, y | s, \alpha_s\right) x_t^i\left(s, y\right) - v^i = -r^i\left(s, \alpha_s\right),$$

and, for all s, i, a^i,

$$\sum_{t,y} \left[p\left(t, y | s, a^i, \alpha_s^{-i}\right) - p\left(t, y | s, \alpha_s\right)\right] x_t^i\left(s, y\right) \le r^i\left(s, \alpha_s\right) - r^i\left(s, a^i, \alpha_s^{-i}\right),$$

as well as, for all T, φ, ψ,

$$\lambda \cdot \sum_{s \in T} x_{\varphi(s)}\left(s, \psi\left(s\right)\right) \le 0.$$

This is a linear program for (x, v). The first set of constraint ensures that α_s yields the same payoff v in all states, the second that playing α_s is a Nash equilibrium, and the third is the same constraint as **(ii)**. Because we assumed that α_s is admissible for all s, the feasible set is non-empty, and because the value of this program is bounded above by $k(\lambda)$, it is finite. We

---

[4]We may think of a Markov decision process (MDP) with irreducible transitions as a stochastic game with a single player and no signal. In this case, take a pure optimal Markov strategy $a^* = (a_s^*)_s$ in the MDP without discounting. The only deviations $(\hat{\alpha}_s)_s$ that satisfy the constraints in the dual $\tilde{\mathcal{P}}(1)$ must improve the objective (they must assign non-positive weights to the actions other than a_s^*, and weight at least one to a_s^*) so that, minimizing over those deviations, it is best to set $\hat{\alpha}_s = a_s^*$ for all s. It follows that $(\beta_s)_s$ is the invariant distribution under the optimal strategy a^*, and the value of the program is equal to the optimal expected (undiscounted) average payoff of the MDP. One can solve the other dual $\tilde{\mathcal{P}}(-1)$ similarly.

shall consider the dual of this linear program. It is

$$\min - \sum_{s,i} \gamma_s^i r^i\left(s, \alpha_s\right) + \sum_{s,i,a^i} \nu_s^i(a^i)\left(r^i\left(s, \alpha_s\right) - r^i\left(s, a^i, \alpha_s^{-i}\right)\right)$$

over $\nu_s^i(a^i) \geq 0, \eta_{T\varphi\psi} \geq 0, \gamma_s^i \in \mathbf{R}$ such that, for all s, t, y, i,

$$p\left(t, y | s, \alpha_s\right) \gamma_s^i - \sum_{a^i} \left[p\left(t, y | s, \alpha_s\right) - p\left(t, y | s, a^i, \alpha_s^{-i}\right)\right] \nu_s^i(a^i) + \lambda^i \sum_{T \supseteq \{s,t\}, \varphi(s) = t, \psi(s) = y} \eta_{T\varphi\psi} = 0,$$

and

$$\lambda^i = -\sum_s \gamma_s^i.$$

There is no loss in assuming $\lambda^i \neq 0$ for all i (we focus on the relevant subset of players otherwise). Define then $\beta_s^i := -\gamma_s^i/\lambda^i$ and $\xi_s^i(a^i) := \nu_s^i(a^i)/\lambda^i$. We get

$$\min \sum_{s,i} \lambda^i \left[\beta_s^i r^i\left(s, \alpha_s\right) + \sum_{a^i} \left(r^i\left(s, \alpha_s\right) - r^i\left(s, a^i, \alpha_s^{-i}\right)\right) \xi_s^i(a^i)\right]$$

over $\eta_{T\varphi\psi} \geq 0, \xi_s^i(a^i)$ with $\xi_s^i(a^i)\mathrm{sgn}\left(\lambda^i\right) \geq 0, \beta_s^i \in \mathbf{R}$ such that $\sum_s \beta_s^i = 1$ for all s, i, such that, for all s, t, y, i,

$$\beta_s^i p\left(t, y | s, \alpha_s\right) + \sum_{a^i} \left(p\left(t, y | s, \alpha_s\right) - p\left(t, y | s, a^i, \alpha_s^{-i}\right)\right) \xi_s^i(a^i) = \sum_{T \supseteq \{s,t\}, \varphi(s) = t, \psi(s) = y} \eta_{T\varphi\psi}.$$

Note that, taking the sum over (t, y), we have

$$\beta_s^i = \sum_{T \ni s, \varphi, \psi} \eta_{T\varphi\psi},$$

and so $\beta_s^i =: \beta_s$ is nonnegative and independent of i. Furthermore, by adding over s, we get that $\sum_{T, \varphi, \psi} |T| \eta_{T\varphi\psi} = 1$. Note also that, if $\beta_s = 0$ for some s, then $\sum_{T \ni s, \varphi, \psi} \eta_{T\varphi\psi} = 0$, and so, because $\eta_{T\varphi\psi} \geq 0$, it follows that

$$\sum_{a^i} \left(p\left(t, y | s, \alpha_s\right) - p\left(t, y | s, a^i, \alpha_s^{-i}\right)\right) \xi_s^i(a^i) = 0;$$

Furthermore, because, given s and i, the variables $\xi_s^i(a^i)$ all have the same sign independently of a^i, either $\sum_{a^i} \xi_s^i(a^i) \neq 0$, or $\xi_s^i(a^i) = 0$ for all a^i. Note that, in the former case, we can define the strategy $\tilde{\alpha}_s^i \in \Delta(A^i)$ by $\tilde{\alpha}_s^i(a^i) = \xi_s^i(a^i)/\sum_{\tilde{a}^i} \xi_s^i(\tilde{a}^i)$, and admissibility then implies that the corresponding term in the objective function is nonnegative, and setting $\xi_s^i(a^i) = 0$ for all a^i would achieve a value at least as low. Hence, if $\beta_s = 0$ for some s, we can assume $\xi_s^i(a^i) = 0$ for all i, a^i, and the terms in the objective and the constraints that involve the state s vanish. Therefore, we might as well assume $\beta_s > 0$ for all s.

Given ξ_s^i, we define $\hat{\alpha}_s^i \in \mathbf{R}^{A^i}$ by, for all a^i,

$$\hat{\alpha}_s^i\left(a^i\right) = \alpha_s^i\left(a^i\right) + \frac{\alpha_s^i\left(a^i\right)}{\beta_s} \sum_{\tilde{a}^i} \xi_s^i(\tilde{a}^i) - \frac{\xi_s^i(a^i)}{\beta_s}.$$

Note that for all s, i, $\sum_{a^i} \hat{\alpha}_s^i\left(a^i\right) = 1$ for all s, i, $\hat{\alpha}_s^i(a^i) \leq 0$ if $\lambda^i > 0$ and $\alpha_s^i(a^i) = 0$, and $\hat{\alpha}_s^i(a^i) \geq 0$ if $\lambda^i < 0$ and $\alpha_s^i(a^i) = 0$. Conversely, given such $\hat{\alpha}_s^i$, we can set

$$\xi_s^i(a^i) = M\alpha_s^i(a^i)\mathrm{sgn}(\lambda^i) + \beta_s(\alpha_s^i(a^i) - \hat{\alpha}_s^i(a^i))$$

with large M so that $\xi_s^i(a^i)\mathrm{sgn}(\lambda^i) \geq 0$. Thus we can rewrite our problem as

$$\min \sum_{s,i} \lambda^i \beta_s r^i\left(s, \hat{\alpha}_s^i, \alpha_s^{-i}\right),$$

over $(\hat{\alpha}_s^i)_{s,i}$ with $\sum_{a^i} \hat{\alpha}_s^i\left(a^i\right) = 1$, $\hat{\alpha}_s^i(a^i) \leq 0$ if $\lambda^i > 0$ and $\alpha_s^i(a^i) = 0$, and $\hat{\alpha}_s^i(a^i) \geq 0$ if $\lambda^i < 0$ and $\alpha_s^i(a^i) = 0$ as well as $\beta_s \geq 0$, $\sum_s \beta_s = 1$, and $\eta_{T\varphi\psi} \geq 0$, such that, for all s, t, y, i,

$$\beta_s p\left(t, y|s, \hat{\alpha}_s^i, \alpha_s^{-i}\right) = \sum_{T \supseteq \{s,t\}, \varphi(s)=t, \psi(s)=y} \eta_{T\varphi\psi}. \tag{1}$$

Note that if $\beta_s > 0$, then it follows from (1) that $p\left(t, y|s, \hat{\alpha}_s^i, \alpha_s^{-i}\right)$ is nonnegative and independent of i. Also, if $\beta_s = 0$, we can assume $\hat{\alpha}_s^i = \alpha_s^i$ for all i without loss in the objective function. Thus in both cases, we can assume that $\hat{p}(t, y|s) := p(t, y|s, \hat{\alpha}_s^i, \alpha_s^{-i}) \geq 0$. Also note that $(\beta_s)_s$ is an invariant distribution of the transition function $\hat{p}(t \times Y|s)$. To see this, take the sum of (1) over s, y, and we have

$$\sum_s \beta_s \hat{p}(t \times Y|s) := \sum_{T \ni t, \varphi, \psi} \eta_{T\varphi\psi} = \beta_t.$$

Conversely, if $(\beta_s)_s$ is an invariant distribution of $\hat{p}(t \times Y|s)$, then it follows from Lemma 1 of HSTV that there exists $\eta_{T\varphi\psi} \geq 0$ that satisfies (1).[5]

Thus we can rewrite our problem without using $\eta_{T\varphi\psi}$ as follows:

$$\min \sum_{s,i} \lambda^i \beta_s r^i \left(s, \hat{\alpha}_s^i, \alpha_s^{-i}\right),$$

over $(\hat{\alpha}_s^i)_{s,i}$ for all i for which $\lambda^i \neq 0$, with $\sum_{a^i} \hat{\alpha}_s^i(a^i) = 1$, $\hat{\alpha}_s^i(a^i) \leq 0$ if $\lambda^i > 0$ and $\alpha_s^i(a^i) = 0$, and $\hat{\alpha}_s^i(a^i) \geq 0$ if $\lambda^i < 0$ and $\alpha_s^i(a^i) = 0$, and

$$\hat{p}(t, y|s) := p\left(t, y|s, \hat{\alpha}_s^i, \alpha_s^{-i}\right) \geq 0,$$

as well as $\beta_s \geq 0$, $\sum_s \beta_s = 1$ such that $(\beta_s)_s$ is an invariant distribution of $\hat{p}(t \times Y|s)$. (If there are multiple invariant distributions, use the one that minimizes the objective function.) Taking the supremum over admissible $(\alpha_s)_s$, this gives us precisely $\tilde{\mathcal{P}}(\lambda)$.

4   The Structure of Equilibrium Payoffs in Repeated Games

This section focuses on repeated games with imperfect public monitoring. Throughout, the sets of actions and signals are finite, and attention is restricted to perfect public equilibria. For a fixed discount factor $\delta < 1$, $E(\delta)$ denotes the set of mixed-strategy PPE payoffs, and $E^p(\delta)$ denotes the set of pure-strategy PPE payoffs. The limits of these equilibrium payoff sets (as $\delta \to 1$) are denoted by $E = \lim_{\delta \to 1} E(\delta)$ and $E^p = \lim_{\delta \to 1} E^p(\delta)$, respectively. We show that (i) E^p has either empty interior or is a polytope; (ii) the result does not extend to E, which includes mixed-strategy equilibria.

The characterization of Fudenberg and Levine (1994) implies that these limits E and E^p are well-defined, and that E and E^p are compact, convex and semialgebraic by the Tarski-Seidenberg theorem. (The extension by Fudenberg, Levine and Takahashi (2007) establishes that this is true even if these limit sets have empty interior.) In addition, both E and E^p are independent of the availability of a public randomization device. In the absence of such a device, none of these properties (except compactness) holds for a fixed discount factor, as explained below.

Because their program is such that the vector of weights λ appears both in the constraints

---

[5]Use the indicator function of (s,t) for $(x_t(s))$ in the notation of Lemma 1. Note that one can easily generalize Lemma 1 to cases without irreducibility.

and in the objective, it is difficult to obtain sharper results from the primal. In contrast, because the constraints in the dual only involve the signs of the weights, for each admissible action profile $a \in A$ and each orthant of λ, we have finitely many linear constraints on $\hat{\alpha}$, which are independent of λ. The result then follows since there are only finitely many candidates of $\hat{\alpha}$ that can minimize the linear objective.

Corollary 1 *Assume that E^p has non-empty interior. Then E^p is a polytope.*[6]

Before proving this result, let us briefly mention what is known about $E^p(\delta)$ and $E(\delta)$ for fixed $\delta < 1$. Most of the results are for the case of perfect monitoring. For the case in which a public randomization device is allowed, Abreu and Sannikov (2011) show that the equilibrium payoff set $E^p(\delta)$ is a polytope, when there are only two players and monitoring is perfect. Furthermore, they show that the set of vertices is no more than thrice the number of action profiles. It is not known whether their results generalize to mixed strategies, more players or imperfect monitoring (see below, however). If no public randomization device is assumed, neither $E(\delta)$ nor $E^p(\delta)$ need be convex: non-convexity is shown by Sorin (1986), and Yamamoto (2010) provides an example in which this is true for discount factors arbitrarily close to one. More generally, the set of equilibrium payoffs $E(\delta)$ need not be semialgebraic. See Berg and Kitti (2010) for examples of the fractal nature of $E^p(\delta)$, which generalizes easily to mixed strategies for low enough discount factors.

Proof. Since E^p has non-empty interior, by Fudenberg and Levine (1994), Fudenberg, Levine and Takahashi (2007), and our dual characterization applied to pure-strategy equilibria, we have

$$E^p = \bigcap_{\lambda \in \mathbf{R}^I} \{v \in \mathbf{R}^I \mid \lambda \cdot v \le k^p(\lambda)\},$$

where $k^p(\lambda)$ is the solution to the following program:

$$\max_{a \in A, \text{ admissible}} \min_{\hat{\alpha} \in D(a, \text{sgn}(\lambda))} \lambda \cdot r(a, \hat{\alpha}),$$

where $\text{sgn}(\lambda) = (\text{sgn}(\lambda^i))_i$, $r(a, \hat{\alpha}) = (r^i(\hat{\alpha}^i, a^{-i}))_i$, and for each profile of signs $\zeta = (\zeta^i)_i \in \{-1, 0, 1\}^I$, $D(a, \zeta)$ is the set of profiles $\hat{\alpha} = (\hat{\alpha}^i)_i \in \times_{i \in I} \mathbf{R}^{A_i}$ such that for each $i \in I$ with

---

[6]If E^p has empty interior, neither the primal nor the dual program applies. The primal has been generalized by Fudenberg, Levine and Takahashi (2007) to include this case, but we have not explored the dual of their program. Clearly, with two players, the result extends to the case in which E^p has empty interior, as the set E^p must then be either a line segment or a point.

$\zeta^i \neq 0$, $\sum_{\tilde{a}^i \in A^i} \hat{\alpha}^i(\tilde{a}^i) = 1$, $\hat{\alpha}^i(\tilde{a}^i) \leq 0$ if $\zeta^i = 1$ and $\tilde{a}^i \neq a^i$, $\hat{\alpha}^i(\tilde{a}^i) \geq 0$ if $\zeta^i = -1$ and $\tilde{a}^i \neq a^i$, and $p(y|\hat{\alpha}^i, a^{-i}) \geq 0$ is independent of i such that $\zeta^i \neq 0$. Our proof further exploits the following two properties of the dual characterization: (1) $D(a, \text{sgn}(\lambda))$ depends on λ only through the profile of signs of λ^i, and (2) $D(a, \text{sgn}(\lambda))$ is a convex polytope.

The proof will use the following standard results (see Rockafellar, 1970): (i) any polyhedron D admits a finite subset D^* such that any linear function on D is minimized at some point in D^*; (ii) the convex hull of a finite union of polyhedral cones is a polyhedral cone; (iii) the polar cone of a polyhedral cone is a polyhedral cone.[7]

For each $\zeta \in \{-1, 0, 1\}^I$, we define

$$\Lambda(\zeta) = \{\lambda \in \mathbf{R}^I \mid \text{sgn}(\lambda) = \zeta\}.$$

We also define $\bar{\Lambda}(\zeta)$ as the closure of $\Lambda(\zeta)$

$$\bar{\Lambda}(\zeta) = \{\lambda \in \mathbf{R}^I \mid \forall i \in I, \ \text{sgn}(\lambda^i) \in \{0, \zeta^i\}\}.$$

Taking the closure simplifies our exposition by allowing us to use standard results on polyhedra and polyhedral cones, which are defined by weak inequalities.

For each $\zeta \in \{-1, 0, 1\}^I$ and $\lambda \in \bar{\Lambda}(\zeta)$, we have

$$k^p(\lambda) \leq \max_{a \text{ admissible}} \min_{\hat{\alpha} \in D(a, \zeta)} \lambda \cdot r(a, \hat{\alpha})$$

since if $\zeta^i \neq 0$ but $\lambda^i = 0$, then the constraint set $D(a, \text{sgn}(\lambda))$ is less restrictive than $D(a, \zeta)$. Therefore, we have

$$E^p = \bigcap_{\lambda \in \mathbf{R}^I} \{v \in \mathbf{R}^I \mid \lambda \cdot v \leq k^p(\lambda)\}$$

$$= \bigcap_{\zeta \in \{-1,0,1\}^I} \bigcap_{\lambda \in \Lambda(\zeta)} \{v \in \mathbf{R}^I \mid \lambda \cdot v \leq k^p(\lambda)\}$$

$$= \bigcap_{\zeta \in \{-1,0,1\}^I} \bigcap_{\lambda \in \bar{\Lambda}(\zeta)} \{v \in \mathbf{R}^I \mid \lambda \cdot v \leq k^p(\lambda)\}.$$

---

[7]To clarify our terminology, a polyhedron is the intersection of finitely many closed half-spaces, which is generated by finitely many points and directions. A polytope is the convex hull of finitely many points, which is equivalent to a bounded polyhedron. A cone is polyhedral if and only if it is generated by finitely many directions. See Rockafellar (1970, Section 19 and Theorem 19.1).

Since E^p is bounded, it is enough to show that for each $\zeta \in \{-1, 0, 1\}^I$,

$$E^p(\zeta) = \bigcap_{\lambda \in \bar{\Lambda}(\zeta)} \{v \in \mathbf{R}^I \mid \lambda \cdot v \le k^p(\lambda)\}$$

is a polyhedron.

For each admissible $a \in A$ and $\zeta \in \{-1, 0, 1\}^I$, note that $D(a, \zeta)$ is a polyhedron, and hence finitely generated, i.e., there exist finitely many points $\beta_1, \ldots, \beta_m \in \times_{i \in I} \mathbf{R}^{A_i}$ and finitely many directions $\beta_{m+1}, \ldots, \beta_n \in \times_{i \in I} \mathbf{R}^{A_i}$ such that any $\hat{\alpha} \in D(a, \zeta)$ is represented as

$$\hat{\alpha} = \mu_1 \beta_1 + \cdots + \mu_m \beta_m + \mu_{m+1} \beta_{m+1} + \cdots + \mu_n \beta_n$$

with $\mu_1 + \cdots + \mu_m = 1$ and $\mu_1, \ldots, \mu_n \ge 0$. Let $D^*(a, \zeta) = \{\beta_1, \ldots, \beta_m\}$. Then for any $\lambda \in \bar{\Lambda}(\zeta)$, the minimization problem

$$\min_{\hat{\alpha} \in D(a, \zeta)} \lambda \cdot r(a, \hat{\alpha})$$

has a solution in $D^*(a, \zeta)$. Note that $D^*(a, \zeta)$ is finite and depends on λ only through its sign ζ.

For each admissible $a \in A$, $\zeta \in \{-1, 0, 1\}^I$, and $\hat{\alpha} \in D^*(a, \zeta)$, let $\bar{\Lambda}(a, \hat{\alpha}, \zeta)$ be the set of all directions $\lambda \in \bar{\Lambda}(\zeta)$ such that the max-min problem

$$\max_{a \text{ admissible}} \min_{\hat{\alpha} \in D^*(a, \zeta)} \lambda \cdot r(a, \hat{\alpha})$$

is solved at $(a, \hat{\alpha})$. Let $\Phi(\zeta)$ be the set of selections φ of $D^*(\cdot, \zeta)$, i.e., the set of functions that map each admissible action a' to $\varphi(a') \in D^*(a', \zeta)$. Then we have

$$\bar{\Lambda}(a, \hat{\alpha}, \zeta) = \bigcup_{\varphi \in \Phi(\zeta)} \left(\bar{\Lambda}(\zeta) \cap \bigcap_{\hat{\alpha}' \in D^*(a, \zeta)} \{\lambda \in \mathbf{R}^I \mid \lambda \cdot (r(a, \hat{\alpha}) - r(a, \hat{\alpha}')) \le 0\} \right.$$

$$\left. \cap \bigcap_{a' \text{ admissible}} \{\lambda \in \mathbf{R}^I \mid \lambda \cdot (r(a, \hat{\alpha}) - r(a', \varphi(a'))) \ge 0\} \right),$$

hence $\bar{\Lambda}(a, \hat{\alpha}, \zeta)$ is a finite union of polyhedral cones, and thus its convex hull is a polyhedral cone (Rockafellar, 1970, Theorem 19.6).

Thus, for each $\zeta \in \{-1, 0, 1\}^I$,

$$E^p(\zeta) = \bigcap_{\lambda \in \bar{\Lambda}(\zeta)} \{v \in \mathbf{R}^I \mid \lambda \cdot v \leq k^p(\lambda)\}$$

$$= \bigcap_{a \text{ admissible}} \bigcap_{\hat{\alpha} \in D^*(a, \zeta)} \bigcap_{\lambda \in \bar{\Lambda}(a, \hat{\alpha}, \zeta)} \{v \in \mathbf{R}^I \mid \lambda \cdot (v - r(a, \hat{\alpha})) \leq 0\}.$$

Here, for each $(a, \hat{\alpha}, \zeta)$, since the convex hull of $\bar{\Lambda}(a, \hat{\alpha}, \zeta)$ is a polyhedral cone, its polar cone (with vertex $r(a, \hat{\alpha})$), $\bigcap_{\lambda \in \bar{\Lambda}(a, \hat{\alpha}, \zeta)} \{v \mid \lambda \cdot (v - r(a, \hat{\alpha})) \leq 0\}$, is also a polyhedral cone (Rockafellar, 1970, Corollary 19.2.2). Therefore, $E^p(\zeta)$ is a finite intersection of polyhedral cones, which is a polyhedron. ∎

This corollary raises a natural question: does the corollary extend to mixed strategies? This is obviously the case when the assumptions for the folk theorem are satisfied—in particular, when monitoring is perfect. We conclude this section with an example establishing that the answer is negative, even in the case of two players. Consider the following 2×2 game with payoffs

	L	R
U	$5, 9$	$0, -3$
D	$8, -3$	$-5, 1$

and two possible signals $Y = \{\bar{y}, \underline{y}\}$ with

$$p(\bar{y} \mid \cdot) = \quad \begin{array}{c|c|c|} & L & R \\ \hline U & .4 & .5 \\ \hline D & .41 & .01 \\ \hline \end{array} \quad, \quad p(\underline{y} \mid a) = 1 - p(\bar{y} \mid a).$$

Note that each player's minmax payoff is 0, and the unique static Nash equilibrium is $(\frac{1}{4}U + \frac{3}{4}D, \frac{5}{8}L + \frac{3}{8}R)$, which supports payoffs $(\frac{25}{8}, 0)$. In the appendix, we sketch steps to compute E,
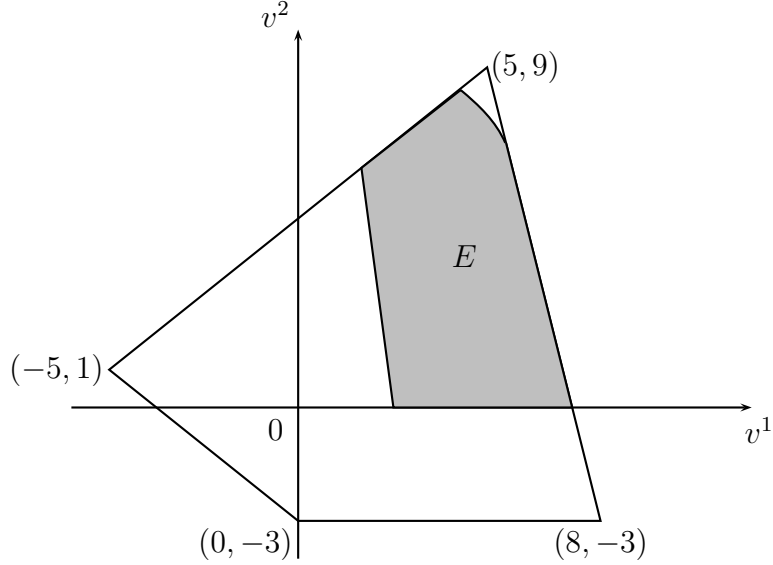
Figure 1: Feasible and limit equilibrium payoff set (shaded area) in the example

which is equal to

$$
E = \left\{ (v^1, v^2) \middle| \begin{array}{l} v^2 \geq \max\left(0, \dfrac{459811 - \sqrt{49070057089}}{12642} - \dfrac{260783 + \sqrt{49070057089}}{64500} v^1\right), \\[2ex] v^2 \leq \min\left(\dfrac{4}{5}v^1 + 5, 29 - 4v^1\right), \\[2ex] v^2 \leq \dfrac{(2\sqrt{2(28 - 5v^1)} + \sqrt{2409})^2}{375} \text{ for } v^1 \leq \dfrac{811}{146} \end{array} \right\}.
$$

Clearly, this set is not a polytope. See Figure 1. As $E(\delta) \to E$, this example shows that the bound of Abreu and Sannikov (2011) on the number of extreme points of $E(\delta)$ cannot possibly extend to mixed strategies and imperfect monitoring.

5 Concluding Comments

Theorem 1 easily extends to the case with short-run players, where the supremum in $\tilde{P}(\lambda)$ is taken over all α such that long-run players play admissible actions and short-run players play static best responses. Similarly, Corollary 1 holds if all (both long- and short-run) players play

15

pure strategies, or if all long-run players play pure strategies and for each pure action profile of long-run players, the induced stage game for short-run players has finitely many static Nash equilibria.

Theorem 1 and Corollary 1 also extend to games with unknown payoffs or signal distributions, where Fudenberg and Yamamoto (2010) obtain a primal program à la Fudenberg and Levine (1994) that characterizes the limit set of belief-free (or perfect type-contingently public ex-post) equilibrium payoffs. Their sufficient conditions for a folk theorem obtain immediately. Example 1 of Hörner and Lovo (2009), with two players and two states, is an instance in which the limit belief-free equilibrium payoff set in mixed strategies is not a polytope.

It is also straightforward to adapt Theorem 1 to the characterization of the limit payoff set achieved by perfect communication equilibria for repeated games with imperfect private monitoring, see Tomala (2009).

But our analysis leaves open many questions, among others:

- Is E^p a polytope even when its interior is empty (in the case in which there are more than two players)?

- Is the limit set of pure-strategy equilibrium payoffs also a polytope in the case of finite stochastic games? (The feasible limit payoff set is known to be a polytope.)

- Does the example showing that E need not be a polytope also establish the same result for $E(\delta)$ (with a public randomization device) for high enough discount factors?

- What are the properties of the set of all Nash and sequential (rather than perfect public) equilibrium payoffs and their limits (if they exist) as $\delta \to 1$?

References

[1] Abreu, D., D. Pearce, and E. Stacchetti (1990). "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, **58**, 1041–1063.

[2] Abreu, D. and Y. Sannikov (2011). "An Algorithm for Two Player Repeated Games with Perfect Monitoring," working paper, Economic Theory Center, Princeton.

[3] Berg, K. and M. Kitti (2010). "Equilibrium Paths in Discounted Supergames," working paper, Aalto University.

[4] Cheng, H. (2004). "Analyzing Repeated Games with Imperfect Monitoring: A New Approach," *mimeo*, University of Southern California.

[5] Fan, K. (1956). "Systems of Linear Inequalities," in *Linear Inequalities and Related Systems*, H. W. Kuhn and A. W. Tucker, Ed., Paper 5. Annals of Mathematics Studies, Vol. 38. Princeton: Princeton Univ. Press.

[6] Fudenberg, D. and D. Levine (1994). "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, **62**, 103–135.

[7] Fudenberg, D., D. Levine, and E. Maskin (1994). "The Folk Theorem with Imperfect Public Information," *Econometrica*, **62**, 997–1040.

[8] Fudenberg, D., D. Levine, and S. Takahashi (2007). "Perfect Public Equilibrium When Players Are Patient," *Games and Economic Behavior*, **61**, 27–49.

[9] Fudenberg, D. and Y. Yamamoto (2010). "Repeated Games Where the Payoffs and Monitoring Structure are Unknown," *Econometrica*, **78**, 1673–1710.

[10] Hernández-Lerma, O. and J.B. Lasserre (1999). *Further Topics on Discrete-Time Markov Control Processes*, Springer-Verlag, New York.

[11] Hoffman, A. J. and R. M. Karp (1966). "On Nonterminating Stochastic Games," *Management Science*, **12**, 359–370.

[12] Hörner, J. and S. Lovo (2009). "Belief-Free Equilibria in Games With Incomplete Information," *Econometrica*, **77**, 453–487.

[13] Hörner, J., Sugaya, T., Takahashi S. and N. Vieille (2011). "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica*, **79**, 1277–1318.

[14] Kandori, M., and H. Matsushima (1998). "Private observations, communication and collusion," *Econometrica*, **66**, 627–652.

[15] Rahman, D. and I. Obara (2010). "Mediated Partnerships," *Econometrica*, **78**, 285–308.

[16] Rockafellar, R. T. (1970). *Convex Analysis*, Vol. 28, Princeton Mathematical Series, Princeton University Press, Princeton, NJ.

[17] Sorin, S. (1986). "On Repeated Games with Complete Information," *Mathematics of Operations Research*, **11**, 147–160.

[18] Tomala, T. (2009). "Perfect communication equilibria in repeated games with imperfect monitoring," *Games and Economic Behavior*, **67**, 682–694.

[19] Yamamoto, Y. (2010). "The Use of Public Randomization in Discounted Repeated Games," *International Journal of Game Theory*, **39**, 431–443.

Appendix: Sketch of computations for the example

We use the dual program $\tilde{\mathcal{P}}(\lambda)$ to compute the maximum score $k(\lambda)$ for each direction $\lambda = (\lambda^1, \lambda^2) \in \mathbf{R}^2$. In particular, we analyze 8 cases $(\lambda^1, \lambda^2 \gtrless 0)$ separately.

In the case of $\lambda = (1, 0)$, we achieve $k(1, 0) = 8$ by enforcing $\alpha = (D, L)$.

In the case of $\lambda^1 > 0$ and $\lambda^2 = 1$, we achieve (i) $k(\lambda^1, 1) = 5\lambda^1 + 9$ by enforcing $\alpha = (U, L)$ if $0 < \lambda^1 \leq \frac{2}{5}$, (ii)

$$k(\lambda^1, 1) = \frac{420(\lambda^1)^2 + 437\lambda^1}{75\lambda^1 - 8}$$

by enforcing

$$\alpha = \left(\frac{60\lambda^1 - 2}{75\lambda^1 - 8} U + \frac{15\lambda^1 - 6}{75\lambda^1 - 8} D, L \right)$$

if $\frac{2}{5} < \lambda^1 < 4$, and (iii) $k(\lambda^1, 1) = 8\lambda^1 - 3$ by enforcing $\alpha = (D, L)$ if $\lambda^1 \geq 4$.

In the case of $\lambda = (0, 1)$, we achieve $k(0, 1) = 9$ by enforcing $\alpha = (U, L)$.

In the case of $\lambda^1 < 0$ and $\lambda^2 = 1$, we achieve (i) $k(\lambda^1, 1) = 5\lambda^1 + 9$ by enforcing $\alpha = (U, L)$ if $-\frac{4}{5} \leq \lambda^1 < 0$ or (D, R), and (ii) $k(\lambda^1, 1) = -5\lambda^1 + 1$ by enforcing $\alpha = (D, R)$ if $\lambda^1 < -\frac{4}{5}$.

In the case of $\lambda = (-1, 0)$, we achieve $k(-1, 0) = 0$ by enforcing $\alpha = (U, R)$.

In the case of $\lambda^1 < 0$ and $\lambda^2 = -1$, the south-west border of E is driven by $\lambda^1 \approx -7.5$, where the maximal score is achieved by enforcing either $\alpha = (U, R)$, which attains $k((\lambda^1, -1), (U, R)) = -\frac{250}{49}\lambda^1 - 57$, or $\alpha = (D, pL + (1 - p)R)$ with

$$p = \frac{-50\lambda^1}{-80\lambda^1 - 50},$$

which attains

$$k((\lambda^1, -1), (D, pL + (1 - p)R)) = -\frac{250(\lambda^1)^2 + 507\lambda^1 + 146}{-80\lambda^1 - 50}.$$

By equating the two values of $k((\lambda^1, -1), \alpha)$, we have

$$\lambda^1 = -\frac{260783 + \sqrt{49070057089}}{64500}$$

and
$$k(\lambda^1, -1) = -\frac{459811 - \sqrt{49070057089}}{12642}.$$

This gives the constraint

$$v^2 \geq \frac{459811 - \sqrt{49070057089}}{12642} - \frac{260783 + \sqrt{49070057089}}{64500} v^1.$$

In the case of $\lambda = (0, -1)$, we achieve $k(0, -1) = 0$ by enforcing $\alpha^1 = \frac{1}{4}U + \frac{3}{4}D$ and any $\alpha^2 \neq \frac{49}{50}L + \frac{1}{50}R$.

In the case of $\lambda^1 > 0$ and $\lambda^2 = -1$, we achieve $k(\lambda^1, -1) = 8\lambda^1 + 3$ by enforcing $\alpha = (D, L)$.

Summarizing those computations, we obtain $E = \{v \in \mathbf{R}^2 \mid \lambda \cdot v \leq k(\lambda) \, \forall \lambda \in \mathbf{R}^2\}$. In particular, the north-east boundary of E contains a smooth boundary:

$$v^2 = \frac{(2\sqrt{2(28 - 5v^1)} + \sqrt{2409})^2}{375}$$

for $\frac{325 + 54\sqrt{1606}}{578} \leq v^1 \leq \frac{811}{146}$.