A "Pencil-Sharpening" Algorithm for Two Player Stochastic Games with Perfect Monitoring*

Dilip Abreu Benjamin Brooks Yuliy Sannikov

April 28, 2016

Abstract

We study the subgame perfect equilibria of two player stochastic games with perfect monitoring and geometric discounting. A novel algorithm is developed for calculating the discounted payoffs that can be attained in equilibrium. This algorithm generates a sequence of tuples of payoffs vectors, one payoff for each state, that move around the equilibrium payoff sets in a clockwise manner. The trajectory of these "pivot" payoffs asymptotically traces the boundary of the equilibrium payoff correspondence. We also provide an implementation of our algorithm, and preliminary simulations indicate that it is more efficient than existing methods. The theoretical results that underlie the algorithm also yield a bound on the number of extremal equilibrium payoffs.

Keywords: Stochastic game, perfect monitoring, algorithm, computation.
JEL classification: C63, C72, C73, D90.

*Abreu: Department of Economics, Princeton University, dabreu@princeton.edu; Brooks: Becker Friedman Institute and Department of Economics, University of Chicago, babrooks@uchicago.edu; Sannikov: Department of Economics, Princeton University, sannikov@princeton.edu. This work has benefitted from the comments of seminar participants at the Hebrew University and UT Austin. We have also benefitted from the superb research assistance of Mathieu Cloutier, Moshe Katzwer, and Kai Hao Yang. Finally, we would like to acknowledge financial support from the National Science Foundation.

1 Introduction

This paper develops a new algorithm for computing the subgame perfect equilibrium payoffs of two player stochastic games with perfect monitoring. Specifically, we study the pure strategy equilibria of repeated games with a stochastically evolving state variable that determines which actions are feasible for the players, and, together with the chosen actions, induces the players' flow payoffs. The chosen actions in turn influence the future evolution of the state. This classical structure is used to describe a wide range of phenomena in economics and in other disciplines. The range of applications include: dynamic oligopoly with investment (in, e.g., capacity, research and development, advertising), markets for insurance against income shocks, and the dynamics of political bargaining and compromise (cf. Ericson and Pakes, 1995; Kocherlakota, 1996; Dixit, Grossman, and Gul, 2000).

Our work has three inter-related components: (i) we uncover new theoretical properties of the equilibria that generate extreme payoffs for a fixed discount factor, (ii) we use these properties to develop a new algorithm for calculating the set of all equilibrium payoffs, and (iii) we provide a user-friendly implementation that other researchers can use to specify, solve, and analyze their games of interest. Preliminary results indicate that our algorithm is significantly more efficient than previously known computational procedures.

The standard methodology for characterizing subgame perfect equilibrium payoffs for infinitely repeated games comes from Abreu, Pearce, and Stacchetti (1986, 1990), hereafter APS. They showed that the set of discounted payoffs that can arise in subgame perfect equilibria satisfies a recursive relationship, which is analogous to the Bellman equation from dynamic programming. This recursion stems from the fact that any equilibrium payoff can be decomposed as the flow payoff from one period of play plus the expected discounted payoff from the next period onwards, which, by subgame perfection, is also an equilibrium payoff. Just as the value function is the fixed point of a Bellman operator, so too the equilibrium payoff set is the largest fixed point of a certain set operator, which maps a set of payoffs which can be promised as continuation utilities into a set of new payoffs which they generate. In addition, APS show that iterating this operator on a sufficiently large initial estimate will yield a sequence of approximations that asymptotically converges to the true equilibrium payoff set. Although APS wrote explicitly about games with imperfect monitoring and without a state variable, their results extend in an obvious way to the case of perfect monitoring and a stochastic state whose evolution is influenced by the players' actions.[1]

[1]For early extensions involving a state variable see Atkeson (1991) and Phelan and Stacchetti (2001). A more recent application is Hörner et al. (2011). For a more complete description of the self-generation methodology for stochastic games, see Mailath and Samuelson (2006).

The APS algorithm does not exploit the detailed structure of equilibria nor does it focus attention on equilibria that generate extreme payoffs.[2] In contrast, Abreu and Sannikov (2014), hereafter AS, provide an algorithm that does this for two-player *repeated* games with perfect monitoring, that is, the same environment studied here but without the stochastic state. The algorithm of AS exploits the simple structure of the equilibria that attain extreme payoffs. Some extreme payoffs are generated with both players strictly preferring their first-period action over any deviations, while for other payoffs, at least one player is indifferent to deviating to another action. AS show that in the former case, the corresponding equilibrium involves the repetition of the same action profile in every period. In the latter case, it turns out that there are at most four payoffs that can be used as continuation values.[3] Thus, associated with each action profile there are at most four extremal equilibrium payoffs when incentive constraints bind in the first period and at most one when incentive constraints are slack. This leads to an obvious bound on the number of extreme points when the action sets are finite as AS assume (and as we assume in this paper as well).

With the generalization to a stochastic game, there is not one set of equilibrium payoffs, but rather a set of such payoffs for each possible initial state. In this richer setting, simultaneously considering the generation of a particular *tuple*[4] of payoff vectors, one for each state, leads to computationally useful insights. Consider a tuple of equilibrium payoff vectors that maximize the same weighted sum of players' utilities in every state, e.g., the tuple of equilibrium payoffs that maximize player 1's payoffs in each state. For a generic choice of weights, these maximal payoffs are unique and are in fact extreme points of their respective equilibrium payoff sets. We show that the equilibria that generate such maximal payoff tuples have a great deal of common structure. Specifically, we show that *behavior in these equilibria follows a common stationary structure until the first history at which some player's incentive constraint binds.*

To illustrate, suppose that in the initial period, the state is s and in the extremal equilibrium all players *strictly* prefer their equilibrium actions a_i over any deviation. Then it must be the case that if the state in the second period turns out (by chance) to also be s, exactly

[2]APS do not even assume the existence of a public randomization device, so the set of equilibrium payoffs need not be convex.

[3]Due to there being perfect monitoring and two players, the locus of continuation payoffs that make a given player indifferent to deviating is a line, and the intersection of that binding incentive constraint with the (convex) set of equilibrium payoffs that are incentive compatible for both players has at most two extreme points. There are therefore at most four extreme binding continuation values between the two players' incentive constraints, and it is one of these payoffs which must be generated by continuation play.

[4]Throughout our exposition, we will use the term *tuple* to denote a function whose domain is the set of states and the term payoff will usually refer to a vector specifying a payoff for each player.

the same actions a_i must be played.[5] Moreover, suppose that the state switches to some s' at which incentive constraints are again slack, and then returns to s. Still, the players must use the original actions a_i. It is only after the state visits some s'' at which at least one player is indifferent to deviating that the stationarity property may break, and subsequent visits to s or s' may be accompanied by different equilibrium actions. This stationarity is reminiscent of the classical observation that Markov decision problems admit an optimal policy that is stationary (Blackwell, 1965). Furthermore, as in AS, there are still at most four payoffs that may be generated by continuation play when the actions a are played in the first period and some player is indifferent to deviating.

The tuples of payoffs that are generated by equilibria with this structure, i.e., stationarity until constraints bind and extreme binding continuation values, can be succinctly described by what we refer to as a *basic pair*, which consists of (i) a tuple of pairs of actions that are played in each state in the first period and (ii) a tuple of *continuation regimes* that describe how play proceeds from period two onwards. The continuation regime for a given state either indicates (iia) that an incentive constraint binds for some player and which extreme binding continuation value is used, or (iib) that incentive constraints are slack and that play will be stationary until a binding state is reached. Thus, the continuation values in (iib) are implicitly taken to be the generated tuple of equilibrium payoffs themselves. The basic pair is in a sense a generalization of the familiar decomposition of equilibrium payoffs into a discount-weighted sum of flow utilities and continuation values, except that it also incorporates the exceptional recursive structure that arises in extremal equilibria when incentive constraints are slack. Since there are only finitely many extreme binding continuation values associated with each action, there are only finitely many ways to configure the basic pair. Hence, there is a finite set of *basic equilibrium payoffs* that are generated by basic pairs and are sufficient to maximize payoffs in any direction (that is for some set of weights over players' utilities).

The second central feature of the theory we develop is a novel algorithm for computing the tuple of equilibrium payoff sets $\mathbf{V}(s)$. This algorithm adopts a methodology that is quite different from the earlier results of APS and AS. We construct an infinite sequence of payoff tuples that move around the equilibrium payoff correspondence in a clockwise direction. As the algorithm progresses, these payoffs move closer and closer to the true equilibrium payoffs, and asymptotically trace the boundary of $\mathbf{V}$. To be more specific, our algorithm constructs a sequence of tuples of payoffs $\mathbf{v}^k$, with one payoff $\mathbf{v}^k(s)$ for each state s, which are estimates of the equilibrium payoffs that all maximize in a given common direction. We refer to the payoff tuples along the sequence as *pivots.* It will always be the case that these pivots are

[5] In exceptional cases, it could be that there is an equivalent action $a_i' \neq a_i$ that could also be played, but even in such cases, a_i may be reused without loss.

generous estimates, in the sense that they are higher in their respective directions than the highest equilibrium payoff. In addition, we keep track of an estimate of the basic pair that generates each pivot. In a sense, this equilibrium structure is analogous to a hybrid of AS and Blackwell, in that we "solve out" the stationary features of the equilibrium, but when non-stationarity occurs, we use a coarse approximation of the extreme binding payoffs that can be inductively generated. Our algorithm generates the sequence of pivot payoffs by gradually modifying the approximate basic pair.

This "pivoting" idea is especially fruitful in combination with another insight, that allows us to make changes only in one state at a time as we move around and compute successive pivots. This is possible because of a remarkable property of the equilibrium sets: it is possible to "walk" around the boundary of $\mathbf{V}$ while stepping only on payoffs that are generated by basic pairs, where each basic pair differs from its predecessor in at most one state. We remark that this property is far from obvious, given the complex synergies that can arise between actions in different states through their effect on transitions. For example, switching actions in state s may lead to a higher probability of state s', which is an unfavorable change unless actions are also changed in s' to facilitate transition to a third state s'' that has desirable flow payoffs. It turns out, however, that one does not need to exploit these synergies when starting from maximal payoffs and moving incrementally along the frontier. Moreover, we show that there is a simple procedure for identifying the direction in which a particular modification will cause payoffs to move. Thus, starting from some incumbent basic pair, there is a straightforward series of calculations that identify the modification that moves payoffs in as "shallow" a direction as possible, and by iteratively computing shallowest directions and making single-state substitutions, one can construct a sequence of basic pairs whose corresponding payoffs demarcate the extent of the equilibrium payoff sets.[6]

This structure is the second pillar of our approach and underlies the algorithm we propose. In particular, our algorithm constructs an analogous sequence of pivots, except that instead of using true basic pairs (which require precise knowledge of equilibrium payoffs), the algorithm constructs approximate basic pairs that use an approximation of the equilibrium payoff correspondence to compute incentive constraints and binding continuation values. At each iteration, the algorithm generates "test directions" that indicate how payoffs would move for every possible modification of the action pair or continuation regime in one of the states. The algorithm identifies the shallowest of these test directions and introduces the corresponding modification into the basic pair. We show that this shallowest substitution will cause the

[6]Strictly speaking, at each step, our algorithms identify an initial substitution in a new action pair and/or continuation regime in single state. A Bellman-like procedure is used to obtain the next pivot. This updating procedure entails no further changes to actions, but the regimes in some states may change from non-binding to binding, in order to preserve incentive compatibility.

pivot to move around the equilibrium payoff correspondence in a clockwise direction but always stay weakly outside. This is the algorithmic analogue of the equilibrium property. As the algorithm progresses, the pivot revolves around and around $\mathbf{V}$. Moreover, every time the pivot moves, we remove payoffs from the approximation that are "above" the trajectory of the pivot. Thus, as the trajectory gets closer to $\mathbf{V}$, the estimate improves, and the pivots move even closer. In the limit, the pivot traces the frontier of the true equilibrium payoff correspondence.

At a high level, the progression of these approximations towards the equilibrium payoff correspondence resembles the process by which a prism pencil sharpener gradually shaves material from a new pencil in order to achieve a conical shape capped by a graphite tip. In this analogy, the final cone represents the equilibrium payoffs and the initial wooden casing represents the extra non-equilibrium payoffs contained in the initial approximation. Every time the pivot moves, it "shaves" off a slice of the excess material. The rotations continue until the ideal conical shape is attained. To complete the analogy it should be noted that a tuple of pencils (one for each state) is being sharpened simultaneously, and in a synchronized way.

As with the APS algorithm, the set of available continuation values starts out large and is progressively refined, as we require the available continuation values to be inductively generated. A key difference is that APS implicitly generates new payoffs using all kinds of equilibrium structures. Here we show that only particular equilibrium structures can generate extreme payoffs. By focusing attention on only those equilibrium structures that are candidates to generate extreme payoffs, the algorithm saves time and other computational resources.

Our procedure is quite different from previous methods for computing equilibrium payoffs, and the complete description requires the introduction of a number of new concepts. We will therefore build slowly towards a general algorithm by first considering the simpler problem of computing the correspondence of feasible payoffs that can arise for some sequence of actions. Our methodology yields a simple "pencil-sharpening" algorithm for calculating this object, and the exposition of this algorithm, in Section 3, allows us to develop intuition and ideas that will be used in the computation of equilibrium payoffs.

In addition to our theoretical results, we have also implemented our algorithm as a software package that is freely available through the authors' website.[7] This package consists of a set of routines that compute the equilibrium payoff correspondence, as well as a graphical interface that can be used to specify games and visualize their solutions. The implementation is standalone, and does not require any third-party software to use. We have used this

[7] www.benjaminbrooks.net/software.shtml

program to explore a number of numerical examples, and we will report computations of the equilibria of risk-sharing games à la Kocherlakota (1996). We also report runtime comparisons with an implementation of the algorithm proposed by Judd, Yeltekin, and Conklin (2003), appropriately adapted to stochastic games.[8] Preliminary simulations indicate that our algorithm can be significantly faster than that of Judd et al.

The rest of this paper is organized as follows. Section 2 describes the basic model and background material on subgame perfect equilibria of stochastic games. Section 3 provides a simple algorithm for calculating the feasible payoff correspondence, to build intuition for the subsequent equilibrium analysis. Section 4 gives our characterization of the equilibria that generate extreme payoffs and explains how one might trace the frontier of the equilibrium payoff correspondence. These insights are used in Section 5 to construct the algorithm for computing equilibrium payoffs. Section 6 presents the risk sharing example, and Section 7 concludes. All omitted proofs are in the Appendix.

2 Setting and background

We study stochastic games in which two players $i = 1, 2$ interact over infinitely many periods. In each period, player i takes an action a_i in a finite set of feasible actions $\mathbf{A}_i(s)$, where s is the current state and lies in a finite set S. We denote by $\mathbf{A}(s) = \mathbf{A}_1(s) \times \mathbf{A}_2(s)$ the set of all action pairs that are feasible in state s. Players receive flow utilities $g_i(a|s)$ when the state is s and actions a are played. In addition, the next period's state s' is drawn from the probability distribution $\pi(s'|a, s)$. Players discount future payoffs at the common rate $\delta \in (0, 1)$. The players' actions and the state of the world are all perfectly observable.

Throughout the following exposition, we will take the pair of actions a to be a sufficient statistic for the state, and simply write $g_i(a)$ and $\pi(s'|a)$.[9] In addition, we will use bold-face to denote functions whose domain is the set of states. Correspondences that map states into sets are denoted by bold upper-case, e.g., $\mathbf{A}$ or $\mathbf{X}$, and functions that map states into actions, scalars, or vectors will generally be denoted by bold lower case, e.g., $\mathbf{a}$ or $\mathbf{x}$. We will abuse notation slightly by writing $\mathbf{x} \in \mathbf{X}$ when $\mathbf{x}(s) \in \mathbf{X}(s)$ for all s.

We will study the equilibrium payoff correspondence $\mathbf{V}$, which associates to each state of the world a compact and convex set of equilibrium payoffs $\mathbf{V}(s) \subset \mathbb{R}^2$ that can be achieved in some pure strategy subgame perfect Nash equilibrium with public randomization, when

[8] The algorithm of Judd, Yeltekin, and Conklin (2003) was originally written for non-stochastic games, but the ideas extend readily to games with a stochastic state variable. This extension has been described by Yeltekin, Cai, and Judd (2015).

[9] This is without loss of generality, since we could simply redefine an action to be the ordered pair (a_i, s) when $a_i \in \mathbf{A}_i(s)$ so that a given action (a_i, s) appears in only one state.

the initial state of the world is s. For a formal definition of an equilibrium in this setting, see Mailath and Samuelson (2006).[10] The techniques of APS can be used to show that $\mathbf{V}$ is the largest bounded self-generating correspondence (cf. Atkeson, 1991; Phelan and Stacchetti, 2001; Mailath and Samuelson, 2006; Hörner et al., 2011). This recursive characterization says that any equilibrium payoff can be decomposed into the sum of (i) a flow payoff which is obtained in the first period and (ii) expected discounted continuation equilibrium payoffs from the second period onwards. Specifically, let $v \in \mathbf{V}(s)$ be generated by a pure strategy a in the first period, and let $\mathbf{w}(s')$ denote the payoff generated by the continuation equilibrium if the state is s' in the second period. Since these continuation values are equilibrium payoffs, we must have $\mathbf{w} \in \mathbf{V}$. Moreover, v and $\mathbf{w}$ must satisfy the *promise keeping* relationship:

$$v = (1-\delta)g(a) + \delta \sum_{s' \in S} \pi(s'|a)\mathbf{w}(s'). \tag{PK}$$

In addition, since v is an equilibrium payoff, neither player must have an incentive to deviate in the first period (incentive constraints after the first period are implicitly satisfied, since $\mathbf{w}(s')$ is an equilibrium payoff for all s'). Since actions are perfectly observable, the continuation payoffs after deviations do not affect the equilibrium payoff, and we may assume without loss of generality that a deviator receives their lowest possible equilibrium payoff in the continuation game. This *equilibrium threat point* is defined by

$$\underline{\mathbf{v}}_i(s) = \min \{w_i | (w_1, w_2) \in \mathbf{V}(s) \text{ for some } w_j\}.$$

That is, $\underline{\mathbf{v}}(s)$ is a vector of the worst equilibrium payoffs for each player in state s. The incentive constraint is therefore that

$$v_i \geq (1-\delta)g(a_i', a_{-i}) + \delta \sum_{s' \in S} \pi(s'|a_i', a_j)\, \underline{\mathbf{v}}_i(s')$$

for all $a_i' \in \mathbf{A}_i(s)$. Rearranging terms, we can write this condition as

$$\sum_{s' \in S} \pi(s'|a)\mathbf{w}_i(s') \geq h_i(a) \tag{IC}$$

[10] Strictly speaking, the definition of an equilibrium in Mailath and Samuelson (2006) differs slightly from the one which we are implicitly using. They assume that there is a probability distribution over the state in the initial period, while we are implicitly assuming that an equilibrium is defined *conditional* on a given initial state.

where

$$h_i(a) = \max_{a_i'} \left[\frac{1-\delta}{\delta} \left(g_i\left(a_i', a_j\right) - g_i(a) \right) + \sum_{s' \in S} \pi\left(s' | a_i', a_j\right) \underline{\mathbf{v}}_i(s') \right].$$

Thus, the function $h(a)$ gives the vector of minimum incentive compatible continuation values that are sufficient to deter deviations from the action pair a.

Since $\mathbf{V}$ is the correspondence of all equilibrium payoffs, every payoff $v \in \text{ext}\mathbf{V}(s)$ for each s must be generated in this manner, using some action pair a in the first period and continuation values drawn from $\mathbf{V}$ itself. The technique of APS is to generalize this recursive relationship in a manner that is analogous to how the Bellman operator generalizes the recursive characterization of the value function in dynamic programming. Explicitly, fix a compact-valued payoff correspondence $\mathbf{W}$. Note that the assumption of compactness of $\mathbf{W}$ is maintained throughout. The associated *threat tuple* is $\underline{\mathbf{w}}(\mathbf{W})(s) \in \mathbb{R}^2$, where

$$\underline{\mathbf{w}}_i(\mathbf{W})(s) = \min \left\{ w_i | (w_1, w_2) \in \mathbf{W}(s) \text{ for some } w_j,\ j \neq i \right\}.$$

For a given action pair $a \in \mathbf{A}(s)$, let

$$h_i(a, \mathbf{W}) = \max_{a_i'} \left[\frac{1-\delta}{\delta} \left(g_i\left(a_i', a_j\right) - g_i(a) \right) + \sum_{s' \in S} \pi\left(s' | a_i', a_j\right) \underline{\mathbf{w}}_i(\mathbf{W})(s') \right].$$

We say that a point v is *generated in state s by the correspondence* $\mathbf{W}$ if there exist $a \in \mathbf{A}(s)$ and $\mathbf{w} \in \mathbf{W}$ such that

$$v = (1-\delta) g(a) + \delta \sum_{s' \in S} \pi(s'|a) \mathbf{w}(s'); \tag{PK$'$}$$

$$\sum_{s' \in S} \pi(s'|a) \mathbf{w}_i(s') \geq h_i(a, \mathbf{W}) \ \forall i = 1, 2. \tag{IC$'$}$$

The correspondence $\mathbf{W}$ is *self-generating* if every $v \in \mathbf{W}(s)$ is a convex combination of payoffs that can be generated in state s by $\mathbf{W}$. In particular, define the operator B by

$$B(\mathbf{W})(s) = \text{co} \left\{ v | v \text{ is generated in state } s \text{ by } \mathbf{W} \right\},$$

where co denotes the convex hull. $\mathbf{W}$ is then self-generating if $\mathbf{W} \subseteq B(\mathbf{W})$ (i.e., $\mathbf{W}(s) \subseteq B(\mathbf{W})(s)$ for all $s \in S$). Note from the definition that this operator is monotonic. Tarski's theorem therefore implies that B has a largest fixed point $\mathbf{V}$, which is in fact the equilibrium payoff correspondence.

In the context of repeated (i.e., non-stochastic) games, APS also propose an iterative procedure for calculating $\mathbf{V}$, which extends naturally to stochastic games as follows: Start with any correspondence $\mathbf{W}^0$ that contains $\mathbf{V}$, and generate the infinite sequence $\mathbf{W}^k = B\left(\mathbf{W}^{k-1}\right)$ for $k \geq 1$. One can show that this sequence converges to $\mathbf{V}$ in the sense that $\cap_{k\geq 0}\mathbf{W}^k = \mathbf{V}$. Moreover, if $\mathbf{W}^0$ is chosen so that $B(\mathbf{W_0}) \subseteq \mathbf{W}_0$, then the correspondences will be monotonically decreasing: $\mathbf{W}^k \subseteq \mathbf{W}^{k-1}$ for all $k > 0$.

Throughout the following, we will assume that a pure strategy equilibrium exists for each possible initial state, so that the sets $\mathbf{V}(s)$ are all non-empty. At the end of Section 5, we will clarify how our algorithm would behave if there are no pure strategy Nash equilibria.

3 Intuition: The feasible payoff correspondence

Before describing our approach to computing $\mathbf{V}$, we will first provide some intuition for our methods by solving a simpler problem: finding the *feasible* payoff correspondence $\mathbf{F}$. $\mathbf{F}(s)$ is defined to be the set of discounted present values generated by all possible distributions over action sequences starting from state s, without regard to incentive constraints, but respecting the transition probabilities between states induced by the actions that are played. In a repeated game, the set of feasible payoffs is just the convex hull of stage-game payoffs, since each action can be played every period. In stochastic games, however, the state variable is changing over time, and a given action cannot be played until its corresponding state is reached. Moreover, the distribution over the sequence of states is determined by the sequence of actions. This simultaneity makes calculating $\mathbf{F}$ a non-trivial task.

For example, consider the simple stochastic game depicted in Table 1. There are two states in which the stage game takes the form of a prisoner's dilemma. For each state, we have written the players' payoffs, followed by the probability of remaining in the same state after playing that action profile. Note that the payoffs in state 2 are equal to the payoffs in state 1, shifted up by the vector $(2, 2)$. In addition, the level of persistence is the same for corresponding action pairs. While it is easy to represent the stage-game payoffs, transition probabilities complicate the choice of best actions overall. For example, if the goal were to maximize the sum of players payoffs, should (C, C) be played in state 2, even though it leads to a lower probability of remaining in the more favorable state 2 than do (C, D) and (D, C)?

We approach the problem as follows. A payoff (vector) $v \in \mathbb{R}^2$ that is an extreme point of $\mathbf{F}(s)$ must maximize some linear objective over all elements of $\mathbf{F}(s)$. In particular, there must exist some vector $d = (d_1, d_2)$ such that the line $\{v + xd \mid x \in \mathbb{R}\}$ is a supporting hyperplane of $\mathbf{F}(s)$ at v. Let us further denote by $\widehat{d} = (-d_2, d_1)$ the counter-clockwise normal to that d, i.e., d rotated 90 degrees counter-clockwise. Supposing that d points clockwise around $\mathbf{F}$,

State 1

	C	D
C	$1,1$ $^{1/3}$	$-1,2$ $^{1/2}$
D	$2,-1$ $^{1/2}$	$0,0$ $^{1/3}$

State 2

	C	D
C	$3,3$ $^{1/3}$	$1,4$ $^{1/2}$
D	$4,1$ $^{1/2}$	$2,2$ $^{1/3}$

Table 1: A simple stochastic game

then it must be the case that v is a solution to

$$\max_{w \in \mathbf{F}(s)} w \cdot \widehat{d}. \tag{1}$$

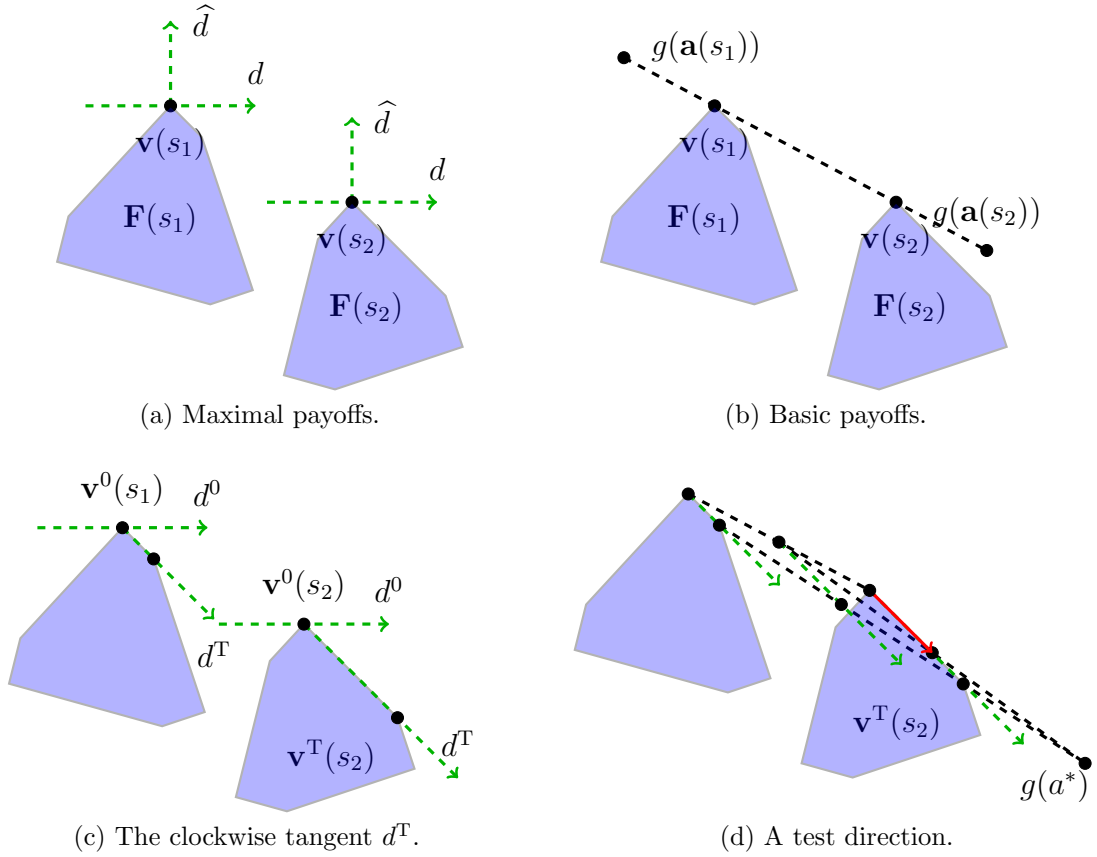
We denote this relationship by saying that the payoff v is *d-maximal*. We may extend this notion to a *tuple* of payoffs $\mathbf{v}$, where $\mathbf{v}(s) \in \mathbf{F}(s)$ for all $s \in S$. If $\mathbf{v}(s)$ is d-maximal for each s, then we will say that the entire tuple of payoffs $\mathbf{v}$ is d-maximal. Finally, payoffs and payoff tuples are maximal if they are d-maximal for some direction d. These definitions are illustrated in Figure 1(a),[11] which depicts a game with two states $S = \{s_1, s_2\}$, with payoffs in state s_1 on the left and payoffs in state s_2 on the right. Highlighted in green are the directions d and its counter-clockwise normal $\widehat{d}$, for which $\mathbf{v}$ is the maximal tuple.

For a fixed vector d, the problem of maximizing $\mathbf{v} \cdot \widehat{d}$ over all feasible distributions over action sequences has the structure of a Markov decision problem, and it is a well-known result in dynamic programming that there exists an optimal solution which is stationary (Blackwell, 1965). In particular, for any direction d there is a d-maximal payoff tuple which is generated by stationary strategies in which the same actions $\mathbf{a}(s)$ are played whenever the state is s. The payoff tuple so generated is the unique solution of the system of equations

$$\mathbf{v}(s) = (1-\delta) g\left(\mathbf{a}(s)\right) + \delta \sum_{s' \in S} \pi\left(s' | \mathbf{a}(s)\right) \mathbf{v}(s') \ \forall s \in S. \tag{2}$$

We refer to payoff tuples that can be generated as the solution to (2) for some choice of action tuple as *basic*. Indeed, every extreme point of $\mathbf{F}(s)$ is part of some basic payoff tuple, and hence can be described as a solution of (2) for some choice of action tuple $\mathbf{a}$. We can see in Figure 1(b) how basic payoff tuples can be generated by a tuple of stationary strategies defined by an action tuple $\mathbf{a}$.

[11] This picture has been drawn for the case in which $\pi\left(s'|a\right) > 0$ for all s' and for all actions which generate extreme payoffs. It is for this reason that all of the edges of $\mathbf{F}\left(s_1\right)$ are parallel to edges of $\mathbf{F}\left(s_2\right)$. More generally, if transition probabilities are degenerate, there does not have to be any particular relationship between the shapes of the feasible payoff sets for states which are not mutually reachable.



(a) Maximal payoffs.

(b) Basic payoffs.

(c) The clockwise tangent d^{T}.

(d) A test direction.

Figure 1: The structure of feasible payoffs.

One could in principle use this observation to characterize the feasible payoff correspondence by solving (2) for *every* possible action tuple to obtain all of the basic payoff tuples and taking the convex hull. An immediate implication is that each $\mathbf{F}(s)$ has a finite number of extreme points. There is, however, a large number of action tuples, and this number grows exponentially in the number of states. One might hope to find an algorithm for characterizing $\mathbf{F}(s)$ where the computational burden depends not on the number of action tuples but rather on the number of actual extreme points, which may be much smaller in practice.

We will argue that such a procedure exists. The naïve algorithm described in the previous paragraph uses the recursive structure of d-maximal payoffs, but there is even more structure to be exploited in how the maximal action tuples change with the direction of maximization. For example, suppose that one had found a maximal action tuple $\mathbf{a}$ with corresponding basic payoff tuple $\mathbf{v}$. A conjecture is that that the action tuples that generate maximal payoff tuples for nearby directions will be similar in structure to $\mathbf{a}$. If this conjecture were correct, we could use the known maximal structure to find other maximal structures by making small modifications to $\mathbf{a}$, and thereby extend our knowledge of the frontier of $\mathbf{F}$.

This intution is indeed correct and is a fundamental building block of the algorithm we propose. Consider a direction d^0, and suppose that we know the actions $\mathbf{a}^0$ which generate a basic payoff tuple $\mathbf{v}^0$ which is d^0-maximal. Let us further suppose that there exists a state in which $\mathbf{F}(s) \neq \{\mathbf{v}(s)\}$. Thus, if the direction of maximization were to rotate clockwise from d^0, $\mathbf{v}^0$ would eventually cease to be maximal in the rotated direction, and there is some critical d^{T} such that if the direction were to rotate any further clockwise, some other extremal tuple would become maximal. The direction d^{T} is in fact the clockwise tangent from $\mathbf{v}^0$ to $\mathbf{V}$. Indeed, if the direction rotated any further, a particular tuple $\mathbf{v}^{\mathrm{T}}$ would become uniquely maximal. These payoffs are *clockwise d^T-maximal*, meaning that of all of the tuples that are d^{T}-maximal, they are the ones that are furthest in the direction d^{T}. Note that at the critical direction d^{T}, both $\mathbf{v}^0$ and $\mathbf{v}^{\mathrm{T}}$ are maximal. Thus, it must be possible to find non-negative scalars $\mathbf{x}(s)$ that are not all zero such that

$$\mathbf{v}^{\mathrm{T}}(s) = \mathbf{v}^0(s) + \mathbf{x}(s)d^{\mathrm{T}}.$$

Figure 1(c) illustrates the directions d^0 and d^{T} in our running example.

We argue that it must be possible to modify $\mathbf{a}^0$ by changing the action pair in a single state, so that the resulting basic payoffs move towards $\mathbf{v}^{\mathrm{T}}$ from $\mathbf{v}^0$. In particular, let s^* be any state in which $\mathbf{x}(s)$ is maximized. In Figure 1, this occurs in state s_2. Since $\mathbf{v}^{\mathrm{T}}$ is uniquely maximal, it must also be basic, and thus $\mathbf{v}^{\mathrm{T}}(s^*)$ is generated by some pure actions a^* in the first period and continuation values $\mathbf{v}^{\mathrm{T}}$, i.e.

$$\mathbf{v}^{\mathrm{T}}(s^*) = (1-\delta)\, g(a^*) + \delta \sum_{s' \in S} \pi(s'|a^*)\, \mathbf{v}^{\mathrm{T}}(s').$$

Now consider the strategy of playing a^* for one period, followed by a return to the original stationary strategies associated with $\mathbf{a}^0$ forever after. The payoff thus generated must be

$$v = (1-\delta) g(a^*) + \delta \sum_{s' \in S} \pi(s'|a^*)\, \mathbf{v}^0(s').$$

As a result, the direction from $\mathbf{v}^0(s^*)$ to v must be

$$\begin{aligned} v - \mathbf{v}^0(s^*) &= \mathbf{v}^{\mathrm{T}}(s^*) - \mathbf{v}^0(s^*) - \delta \sum_{s' \in S} \pi(s'|a^*) \left(\mathbf{v}^{\mathrm{T}}(s') - \mathbf{v}^0(s')\right) \\ &= \left(\mathbf{x}(s^*) - \delta \sum_{s' \in S} \pi(s'|a^*)\, \mathbf{x}(s')\right) d^{\mathrm{T}}, \end{aligned}$$

and since $\mathbf{x}(s^*) \geq \mathbf{x}(s')$ for all s' and $\delta < 1$, it must be that the coefficient on d^{T} is strictly positive. As a result, v must lie in the direction d^{T} relative to $\mathbf{v}^0(s^*)$. Figure 1(d) illustrates the geometry of this construction.

Now, let us consider what would happen if we were to modify $\mathbf{a}^0$ by substituting a^* in place of $\mathbf{a}^0(s^*)$, to obtain a new action tuple $\mathbf{a}^1$, which in turn generate a new basic payoff tuple $\mathbf{v}^1$. We claim that this new payoff tuple $\mathbf{v}^1$ has to lie between $\mathbf{v}^0$ and $\mathbf{v}^{\mathrm{T}}$. To see this, first observe that $\mathbf{v}^1$ must be the fixed point of the following Bellman operator $\mu(\cdot)$:

$$\mu(\mathbf{w})(s) = (1-\delta)g\left(\mathbf{a}^1(s)\right) + \delta \sum_{s' \in S} \pi\left(s'|\mathbf{a}^1(s)\right)\mathbf{w}(s).$$

This operator is a contraction mapping, and so starting from any initial guess of $\mathbf{v}^1$, the iterates will eventually converge to the unique fixed point. Moreover, we have already proven that starting from $\mathbf{v}^0$, $\mu(\mathbf{v}^0)$ moves in the direction d^{T} relative to $\mathbf{v}^0$ (in state s^* only). The linearity of μ in $\mathbf{w}$ implies that subsequent applications of μ will only move payoffs further in the direction d^{T}, as the movement in state s^* is propagated through to the other states. This point is of such importance to our subsequent analysis that we state it as a formal result:

Lemma 1.

(i) For any $\mathbf{a} \in \mathbf{A}$, (2) *has a unique solution.*

(ii) Suppose $\mathbf{v}$ *and* $\widetilde{\mathbf{v}}$ *solve* (2) *for* $\mathbf{a}$ *and* $\widetilde{\mathbf{a}}$ *respectively, that* $\mathbf{v} \neq \widetilde{\mathbf{v}}$, *and that* $\mathbf{a}(s) = \widetilde{\mathbf{a}}(s)$ *for all* $s \neq \widetilde{s}$. *Let*

$$d = (1-\delta)\, g\left(\widetilde{\mathbf{a}}\left(\widetilde{s}\right)\right) + \delta \sum_{s' \in S} \pi\left(s'|\widetilde{\mathbf{a}}\left(\widetilde{s}\right)\right)\mathbf{v}\left(s'\right) - \mathbf{v}\left(\widetilde{s}\right).$$

Then $\widetilde{\mathbf{v}}(s) = \mathbf{v}(s) + \mathbf{x}(s)\, d$ *for some non-negative scalars* $\mathbf{x}(s)$.

Thus, we conclude that there must be at least one action which, when substituted into the initial stationary action sequence, must result in a clockwise movement of the basic payoffs around the feasible correspondence. We should note that generically $\mathbf{v}^1$ is in fact equal to $\mathbf{v}^{\mathrm{T}}$, and only a single substitution will be required to cross each edge of $\mathbf{F}$. In the case of equilibrium studied in Section 4, however, it is a generic possibility that multiple substitutions may be required to move across an edge of $\mathbf{V}$.

Thus far, we have presumed that we omnisciently knew the action a^* that generated the payoffs that were clockwise relative to $\mathbf{v}^0$. This was unnecessary: First, Lemma 1 shows that we can identify the direction in which a substitution moves the payoff tuple by just looking at the first application of μ in the substituted state. In other words, if a new action tuple

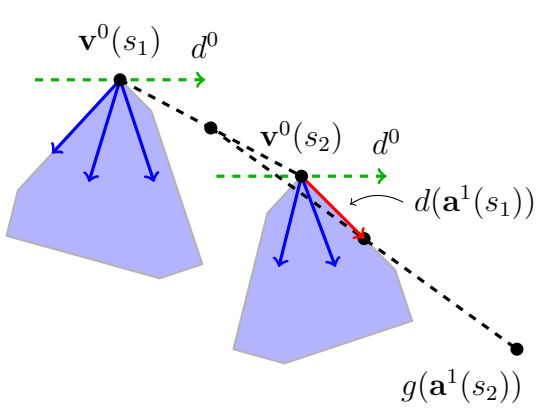
differs from $\mathbf{a}^0$ only in state s in which the action is $a \neq \mathbf{a}^0(s)$, then the direction that the substitution moves payoffs will be

$$d(a) = (1-\delta)g(a) + \delta \sum_{s' \in S} \pi(s'|a)\mathbf{v}^0(s') - \mathbf{v}^0(s).$$

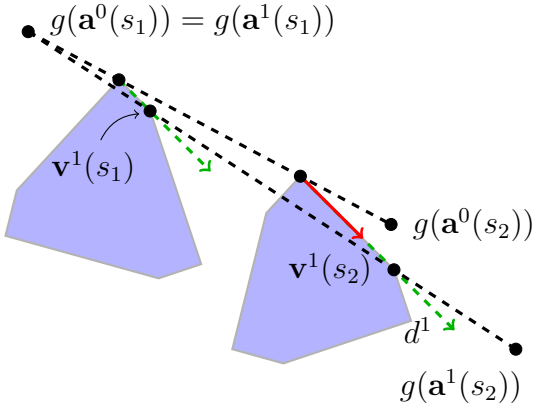
As such, we can easily project where a given substitution will send the payoffs by computing the *test direction* $d\,(a)$. Second, we know that there is *some* substitution that will move us in the direction that points along the frontier. We could therefore consider all substitutions in all states and compare the corresponding test directions. Note that it is impossible for any of the test directions to point above the frontier, since this would imply the existence of a feasible payoff tuple that is outside of $\mathbf{F}$. As a result, the test direction with the smallest clockwise angle of rotation relative to d^0 must point along the frontier, and by implementing the substitution associated with this direction, payoffs are guaranteed to move clockwise along the boundary of $\mathbf{F}$.

From these observations, it follows that there is a simple way, starting from a known $\mathbf{v}^0$, $\mathbf{a}^0$, and d^0, to trace the entire frontier of $\mathbf{F}$ using only basic payoff tuples. We first generate all test directions $d(a)$ for all possible substitutions. One of these directions, denoted by d^1, is *shallowest*, in the sense of having the smallest clockwise angle of rotation from d^0. This shallowest direction must in fact coincide with the tangent direction d^{T}. We then form a new action tuple $\mathbf{a}^1$ by substituting in the action a^* that generated the shallowest test direction and leaving the rest of the actions unchanged, and the new tuple $\mathbf{a}^1$ in turn generates new basic payoffs $\mathbf{v}^1$. We then repeat this process, inductively finding shallowest directions and substituting to generate a sequence of action tuples $\mathbf{a}^k$. These action tuples generate a sequence of payoff tuples $\mathbf{v}^k$ that we refer to as *pivots*, since the direction of movement pivots around these points while tracing the frontier. For generic payoffs, the pivot moves in a new direction each time we make a substitution, though it could in principle move more than once in the same direction. The pivot will eventually visit all of the basic payoff tuples which are uniquely maximal in some direction, at which point we will have traced the frontier of $\mathbf{F}$. Figure 2 demonstrates this "pencil-sharpening" algorithm for feasible payoffs. In this example, the pivot returns to $\mathbf{v}^0$ after six steps.

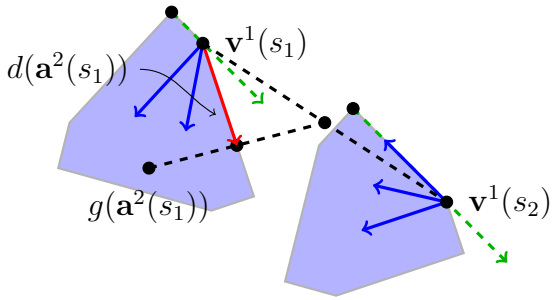
The one remaining difficulty with our proposed algorithm is that it presumes an initial $\mathbf{a}^0$ that generates basic payoffs $\mathbf{v}^0$ that are d^0-maximal. How do we find such a starting point? It turns out that we do not have to! We can augment the game in a simple way that automatically gives us an initial condition. Pick any tuple $\mathbf{v}^0$ that can be weakly separated from $\mathbf{F}$ by some hyperplane. For example, $\mathbf{v}^0(s)$ could be the pointwise maximum of each player's payoffs across all actions in state s, for each s. We use these payoffs to create a tuple



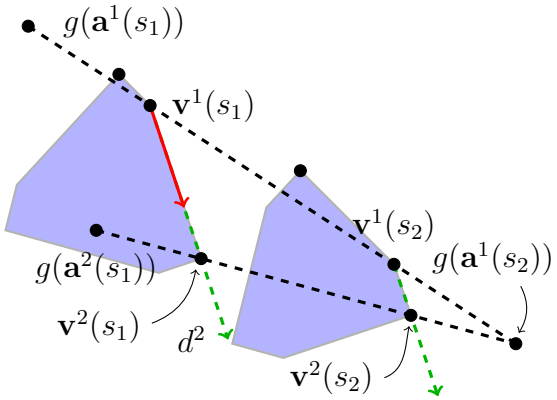
(a) Generate test directions.



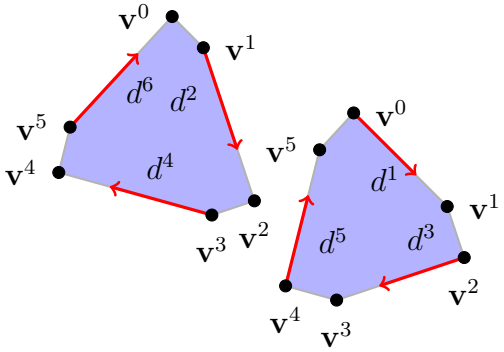
(b) Substitute and advance the pivot.



(c) Generate more test directions.



(d) Substitute again.



(e) Iterate.

Figure 2: Tracing the frontier.

of "synthetic" action pairs $\mathbf{a}^0(s) \notin \mathbf{A}(s)$, and define $g\left(\mathbf{a}^0(s)\right) = \mathbf{v}^0(s)$ and $\pi\left(s|\mathbf{a}^0(s)\right) = 1$, so that $\mathbf{a}^0$ generates the initial pivot. Starting from this initial condition, we trace the boundary of the feasible payoffs of the augmented game which consists of all of the original actions together with the synthetic actions $\mathbf{a}^0$. It is not hard to see that once we pivot around to

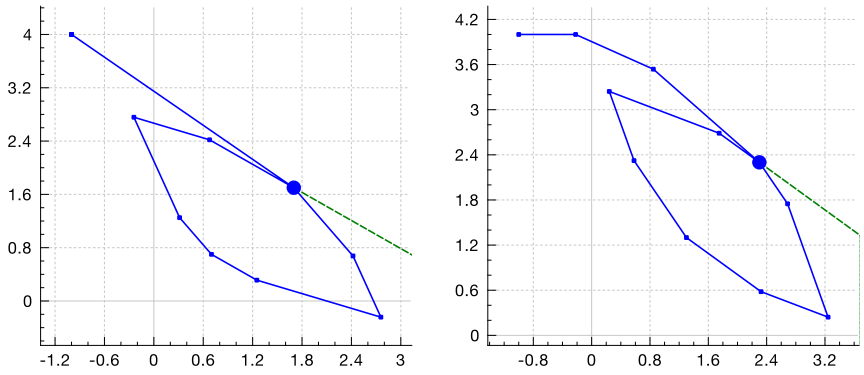


Figure 3: Tracing the boundary of the feasible set for the prisoners' dilemmas example.

the opposite side of $\mathbf{F}$ from $\mathbf{v}^0$, it will not be optimal to use any of the synthetic actions $\mathbf{a}^0$. In fact, these actions will be dominated by *any* choice of actions in the original game. Thus, the synthetic actions will eventually be driven out of the solution, at which point we must be generating payoffs on the boundary of $\mathbf{F}$ for the original game. From that point on, we disallow the synthetic actions $\mathbf{a}^0$ from being substituted back in, and all further pivoting operations will remain on the frontier of $\mathbf{F}$. After one more full revolution, we will have traced the boundary for the original game.

We used this procedure to calculate $\mathbf{F}$ for the game in Table 1. The results of the computation are depicted in Figure 3. The initial pivot consisted of payoffs of $(-1, 4)$ in both states, which corresponds to the lowest payoff for player 1 and the maximum payoff for player 2, across all states and action pairs. After two substitutions, the synthetic actions have been driven out of the system, and we are generating payoffs on the frontier, in particular the symmetric surplus maximizing tuple of payoffs at which (C, C) is played in both states. Note that from the utilitarian efficient payoffs, pivoting from (C, C) to (D, C) in either state shifts the stage payoffs in the direction $(1, -2)$. However, introducing (D, C) in state 2 increases the probability of remaining in the better state 2, whereas introducing (D, C) in state 1 entails a lower probability of state 2. Thus, the correct choice must be to pivot to (D, C) in state 2, and this is indeed the substitution that is identified by our algorithm.

4 The structure of equilibrium payoffs

We now return to the primary objective of this paper, which is the characterization and computation of equilibrium payoffs. The equilibrium requirement adds significant complexity relative to the analysis of the feasible payoffs in the previous section. Equilibrium payoffs are generated by distributions over action sequences that not only obey the transition prob-

abilities between states but also satisfy the players' forward-looking incentive constraints. Nonetheless, we shall see that there are high-level similarities between the structure of feasible payoffs and that of equilibrium payoffs. Characterizing the equilibria that generate extremal equilibrium payoffs is the subject of this section, and in Section 5, we will use these results to develop a general algorithm for computing $\mathbf{V}$.

4.1 Basic pairs

The central tool for characterizing equilibrium payoffs will be what we refer to as a *basic pair*, which in a sense generalizes the stationary strategies of Section 3. This object consists of a tuple of action pairs $\mathbf{a}$ and a tuple of *continuation regimes* $\mathbf{r}$, which we shall explain presently. In Section 2, we reviewed the standard analytical technique of decomposing an equilibrium payoff as the discount-weighted average of a flow payoff and an expected equilibrium continuation value. The basic pair gives this decomposition for an entire *tuple* of equilibrium payoffs $\mathbf{v} \in \mathbf{V}$ simultaneously. In particular, each $\mathbf{v}(s)$ can be decomposed as a weighted average of a flow payoff, $g(\mathbf{a}(s))$, and an expected continuation value w which is determined by $\mathbf{r}(s)$.

The continuation regime, and consequently the expected continuation value, falls into one of two categories depending on whether or not the incentive constraints (IC) are slack or hold with equality. In the *non-binding case*, incentive constraints are slack, $\mathbf{r}(s) = \text{NB}$, and the continuation value is simply the expectation of the payoff tuple $\mathbf{v}$ itself. In the *binding case*, at least one of the players is indifferent to deviating from their prescribed action $\mathbf{a}_i(s)$, so that the expected continuation value lies along a binding incentive constraint. Moreover, this continuation value must be an extreme point of the set of expected equilibrium continuation values at which some constraint binds.

To be more precise, let

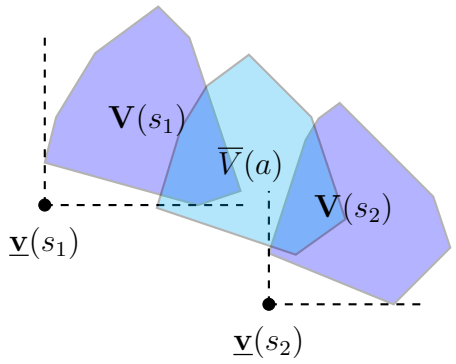
$$\overline{V}(a) = \sum_{s' \in S} \pi(s'|a)\mathbf{V}(s')$$

denote the set of expected equilibrium continuation values when the action pair a is played, and let

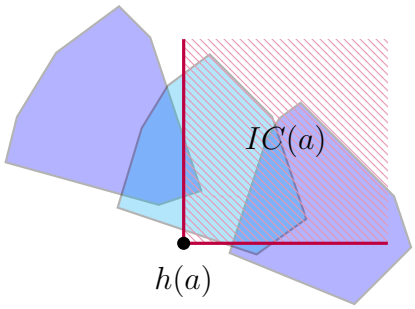
$$IC(a) = \left\{w \in \mathbb{R}^2 | w \geq h(a)\right\}$$

denote the set of continuation value pairs that would deter players from deviating from the action pair a (where the minimal expected continuation value $h(a)$ is defined in Section 2). The set of extreme binding continuation values is

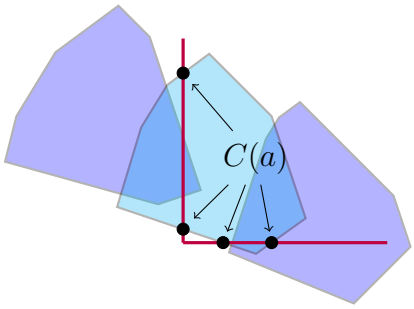
$$C(a) = \text{ext}\left(\overline{V}(a) \cap \text{bd}IC(a)\right),$$



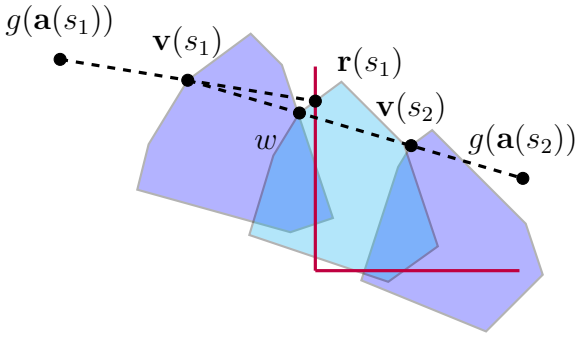
(a) Expected and threat payoffs.



(b) Incentive compatible region.



(c) Extreme binding continuations.



(d) Basic equilibrium payoffs.

Figure 4: The geometry of feasible and incentive compatible continuation values.

where bd denotes the topological boundary, so that bd$IC(a)$ is the set of continuation value vectors at which at least one player is indifferent to deviating.

These sets are depicted for our two-state example in Figure 4. A key observation is that $C(a)$ can have at most four elements. The reason is that bd$IC(a)$ is the union of two rays, so that the intersection of bd$IC(a)$ with $\overline{V}(a)$ is the union of two line segments. Each of these line segments can have at most two extreme points, so that between the two players' constraints there are at most four extreme binding continuation values. Figure 4(c) illustrates the case where $C(a)$ has the maximal number of elements.

Thus, returning to the definition of the continuation regime, either $\mathbf{r}(s) = \text{NB}$ in the non-binding case, or $\mathbf{r}(s) = (\mathbf{B}, w)$ for some $w \in C(\mathbf{a}(s))$ if a constraint binds. As a result, $\mathbf{r}(s)$ can take on at most five values once we have fixed $\mathbf{a}(s)$, so that we can bound the number of basic pairs:

Lemma 2 (Basic pairs)**.** *The number of basic pairs is at most* $5^{|S|}\prod_{s\in S}|\mathbf{A}(s)|$.

We say that the basic pair $(\mathbf{a}, \mathbf{r})$ *generates the payoff tuple* $\mathbf{v}$ if

$$\mathbf{v}(s) = \begin{cases} (1-\delta)\, g(\mathbf{a}(s)) + \delta \sum_{s' \in S} \pi(s'|\mathbf{a}(s))\mathbf{v}(s') & \text{if } \mathbf{r}(s) = \text{NB}; \\ (1-\delta)\, g(\mathbf{a}(s)) + \delta w & \text{if } \mathbf{r}(s) = (\text{B}, w), \end{cases} \tag{3}$$

and if

$$\sum_{s' \in S} \pi(s'|\mathbf{a}(s))\, \mathbf{v}(s') \in IC(\mathbf{a}(s)) \tag{4}$$

whenever $\mathbf{r}(s) = \text{NB}$. In this case, we say that $\mathbf{v}$ is a tuple of *basic equilibrium payoffs*. Equation (3) is a promise keeping condition, analogous to (PK′). Incentive constraints are satisfied by definition when $\mathbf{r}(s) \in C(\mathbf{a}(s))$, and (4) ensures that the expected payoffs themselves are incentive compatible whenever $\mathbf{r}(s) = \text{NB}$. Note that the tuple of payoffs $\mathbf{v}$ that solves (3) is the fixed point of a certain Bellman operator

$$\mu(\mathbf{w}; \mathbf{a}, \mathbf{r})(s) = \begin{cases} (1-\delta)\, g(\mathbf{a}(s)) + \delta \sum_{s' \in S} \pi(s'|\mathbf{a}(s))\mathbf{w}(s') & \text{if } \mathbf{r}(s) = \text{NB}; \\ (1-\delta)\, g(\mathbf{a}(s)) + \delta w & \text{if } \mathbf{r}(s) = (\text{B}, w), \end{cases} \tag{5}$$

Thus, μ maps a tuple of payoffs $\mathbf{w}$ into a new tuple of payoffs $\mu(\mathbf{w}; \mathbf{a}, \mathbf{r})$, where $\mu(\mathbf{w}; \mathbf{a}, \mathbf{r})(s)$ is the discount-weighted average of $g(\mathbf{a}(s))$ and $\mathbf{w}$ when $\mathbf{r}(s) = \text{NB}$. This operator is a contraction of modulus δ, and thus has a unique fixed point which is the solution to (3).

Lemma 3 (Basic equilibrium payoffs)**.** *Suppose that the basic pair* $(\mathbf{a}, \mathbf{r})$ *generates* $\mathbf{v}$*. Then* $\mathbf{v} \in \mathbf{V}$*.*

Proof of Lemma 3. Consider the correspondence $\mathbf{V}'$ defined by $\mathbf{V}'(s) = \mathbf{V}(s) \cup \{\mathbf{v}(s)\}$. It is immediate from the definitions that $\mathbf{V}'$ is self-generating, so that $\mathbf{V}' \subseteq \mathbf{V}$, and hence $\mathbf{v} \in \mathbf{V}$. □

Thus, a basic pair describes the decomposition of an entire tuple of equilibrium payoffs into flows and continuations. Basic pairs correspond to what we might call *basic equilibrium systems*, consisting of one equilibrium for each initial state, that exhibit exceptional recursive structure. In particular, each equilibrium in this system can be decomposed into a flow payoff in the first period and a continuation equilibrium system, which describes the continuation equilibria that are played for each possible state in the second period. A basic equilibrium system has the feature that whenever incentive constraints are slack in the first period, the continuation system simply reboots the original system of equilibria. This corresponds to the case where $\mathbf{r}(s) = \text{NB}$. When $\mathbf{r}(s) \neq \text{NB}$, the continuation equilibrium system generates an extreme binding expected continuation value. This perspective is analogous to how action tuples were used to describe a corresponding tuple of stationary strategies in Section (3). A stationary strategy tuple defines a strategy for each possible initial state, and after the

first period, the strategy tuple simpy restarts. This remains true of the basic equilibrium system when incentive constraints are slack, although when incentive constraints bind, the equilibrium may evolve in a non-stationary manner.

Figure 4(d) gives an example of how basic equilibrium payoffs are generated. In state s_1, on the left, the equilibrium that maximizes player 2' s payoff involves a binding continuation value (at which player 1's incentive constraint is binding), whereas in state s_2, on the right, the equilibrium that maximizes player 2's payoffs has slack constraints. For simplicity, this picture is drawn for the special case where the transition probabilities $\pi(\cdot|\mathbf{a}(s_1))$ and $\pi(\cdot|\mathbf{a}(s_2))$ coincide, so that $\overline{V}(\mathbf{a}(s_1)) = \overline{V}(\mathbf{a}(s_2))$. We suppose, however, that $h(\mathbf{a}(s_1)) > h(\mathbf{a}(s_2))$, so that while the expected pivot w is incentive compatible in state s_2, it is not incentive compatible in s_1.

4.2 The maximality of basic equilibrium payoffs

As we have already indicated, basic pairs have properties which make them extremely useful for the characterization and computation of equilibrium payoffs. Most importantly, it turns out that basic pairs are sufficient to maximize equilibrium payoffs in any given direction. This is in a sense a generalization of the sufficiency of stationary strategies for maximizing feasible payoffs in a given direction.

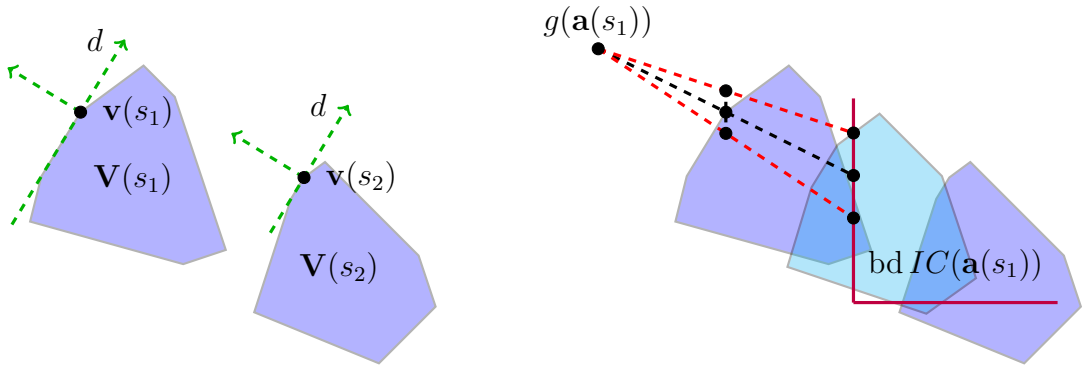
It will be convenient in the sequel to use a refined notion of maximality that selects for a unique maximal payoff tuple. We will say that a payoff tuple $\mathbf{v}$ is clockwise d-maximal in $\mathbf{V}$ if it is d-maximal and if there is no other d-maximal tuple $\mathbf{v}'$ such that $\mathbf{v}'(s) \cdot d > \mathbf{v}(s) \cdot d$ for some s. In other words, among all d-maximal tuples, $\mathbf{v}$ is the one that is furthest in the direction d. Note that while there may be many d-maximal equilibrium payoff tuples, the clockwise d-maximal tuple is unique. Moreover, all of the $\mathbf{v}(s)$ must be extreme points of their respective $\mathbf{V}(s)$. This relationship is depicted in Figure 5(a). In both states s_i , there is a continuum of d-maximal payoffs but a unique clockwise d-maximal payoff, which is $\mathbf{v}(s_i)$.

We have the following result:

Proposition 1 (Maximality of basic equilibrium payoffs)**.** *For every direction d, the clockwise d-maximal equilibrium payoffs are basic.*

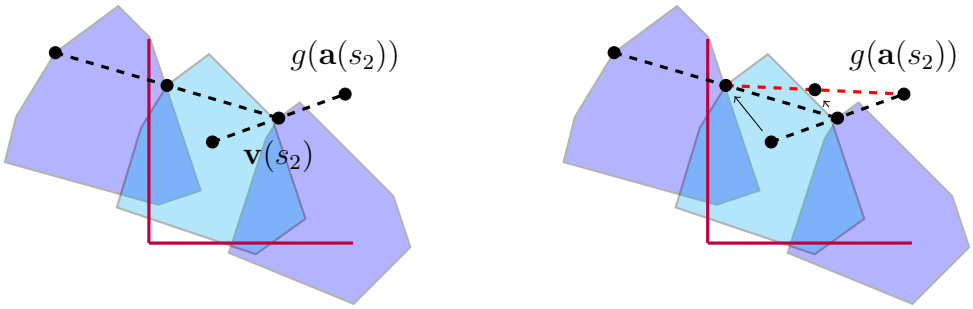
Proof of Proposition 1. Suppose $\mathbf{v}$ is clockwise d-maximal. Since $\mathbf{v}(s) \in \text{ext}\mathbf{V}(s)$, $\mathbf{v}(s)$ must be generated by some pure action $\mathbf{a}(s)$ and expected continuation value $\mathbf{w}(s) \in \overline{V}(\mathbf{a}(s))$. Note that an arbitrary $v \in \mathbf{V}(s)$ may require public randomization over actions in the first period, but this is not true of the extreme points of $\mathbf{V}(s)$.

If (IC) is binding, then it must be that the expected continuation value $\mathbf{w}(s)$ is in $C(\mathbf{a}(s))$. If not, then there must exist a perturbation $\widetilde{d}$ such that $\mathbf{w}(s) + \widetilde{d}$ and $\mathbf{w}(s) - \widetilde{d}$ are both



(a) Clockwise maximal payoffs.

(b) Binding perturbation.



(c) Sub-optimal continuation payoffs.

(d) Non-binding perturbation.

Figure 5: Maximal basic equilibrium payoffs.

feasible and incentive compatible continuation values in $\overline{V}(\mathbf{a}(s)) \cap \mathrm{bd} IC(\mathbf{a}(s))$, so that we can generate the payoffs $\mathbf{v}(s) + \delta\widetilde{d}$ and $\mathbf{v}(s) - \delta\widetilde{d}$, thus contradicting the extremeness of $\mathbf{v}(s)$. This is depicted in Figure 5(b). Thus, $\mathbf{w}(s) \in C(\mathbf{a}(s))$, and we can set $\mathbf{r}(s) = \mathbf{w}(s)$.

On the other hand, suppose that (IC) is slack for both $i = 1, 2$, and

$$\mathbf{w}(s) \neq \sum_{s' \in S} \pi(s'|\mathbf{a}(s))\mathbf{v}(s') = w.$$

This configuration is depicted in Figure 5(c). Note that we must be able to find a $\widetilde{\mathbf{v}}$ such that

$$\mathbf{w}(s) = \sum_{s' \in S} \pi(s'|\mathbf{a}(s))\widetilde{\mathbf{v}}(s'),$$

and since $\mathbf{w}(s) \neq w$, there must be at least one state s'' with $\pi(s''|\mathbf{a}(s)) > 0$ such that $\widetilde{\mathbf{v}}(s'') \neq \mathbf{v}(s'')$. Since the clockwise d-maximal payoff is unique, $\mathbf{v}(s'')$ is either higher in the $\widehat{d}$ direction or in the d direction relative to $\widetilde{\mathbf{v}}(s'')$.

Since $\mathbf{V}$ is convex (because of public randomization), the payoff

$$\widetilde{w} = \sum_{s' \in S \setminus \{s''\}} \pi(s'|\mathbf{a}(s))\widetilde{\mathbf{v}}(s') + \pi(s''|\mathbf{a}(s))((1-\epsilon)\widetilde{\mathbf{v}}(s'') + \epsilon \mathbf{v}(s''))$$

is in $\overline{V}(\mathbf{a}(s))$ for every $\epsilon \in (0,1)$, and since (IC) is slack, there must be a sufficiently small but positive ϵ so that constraints will still be statisfied, i.e., $\widetilde{w} \geq h(a)$ (the constraint is satisfied strictly at $\epsilon = 0$). Thus, it is possible to generate the payoff

$$\begin{aligned}\widetilde{v} &= (1-\delta)g(\mathbf{a}(s)) + \delta\widetilde{w} \\ &= \mathbf{v}(s) + \delta\pi(s''|\mathbf{a}(s))\epsilon(\mathbf{v}(s'') - \widetilde{\mathbf{v}}(s'')).\end{aligned}$$

We can see the construction of this payoff in Figure 5(d). Thus, $\widetilde{v}$ must be higher in the $\widehat{d}$ or d direction, thus contradicting clockwise d-maximality of $\mathbf{v}(s)$.

As a result, it must be that $\widetilde{\mathbf{v}}(s') = \mathbf{v}(s')$ for states in which $\pi(s'|\mathbf{a}(s)) > 0$, and it is obviously without loss of generality to take this to be true when $\pi(s'|\mathbf{a}(s)) = 0$. We can then set $\mathbf{r}(s) = \text{NB}$, so that $\mathbf{v}(s)$ is a solution to (3) and must therefore be a basic equilibrium payoff. □

Intuitively, if incentive constraints are slack and the continuation values are not $\mathbf{v}$, then it is possible to move the continuation payoffs in the direction of $\mathbf{v}$ without violating incentive constraints or feasibility. Since $\mathbf{v}$ is already presumed to be clockwise d-maximal, the continuation values move weakly in the directions $\widehat{d}$ and d, and strictly in at least one of these directions. This means that the payoffs that we generate with these continuation values have also moved in the direction $\widehat{d}$ or d relative to $\mathbf{v}$, which would violate our hypothesis that $\mathbf{v}$ is already clockwise d-maximal.[12]

Since every extremal equilibrium payoff is clockwise maximal for some direction, Proposition 1 implies that it must be generated by a basic pair. Combining this observation with Lemma 2, we have the following result:

Corollary 1 (Number of extremal equilibrium payoffs)**.** *For each s, the number of extreme points of $\mathbf{V}(s)$ is at most $5^{|S|}\prod_{s' \in S}|\mathbf{A}(s')|$.*

[12]This result has some antecedents in the literature. For example, Kocherlakota (1996) studies the Pareto efficient equilibria of a model of informal insurance, and he shows that ratios of marginal utilities of agents should be held constant over time when incentive constraints are slack. Ligon, Thomas, and Worrall (2000, 2002) make a similar point in the context of a more general class of insurance games. Dixit, Grossman, and Gul (2000) make similar observations in the context of a model of political power-sharing. These results are implied by the sufficiency of basic pairs for generating extremal payoffs, as basic pairs build in the property that as long as incentive constraints do not bind, continuation payoffs must maximize a fixed weighted sum of players' utilities.

This result, like a similar one in AS for the non-stochastic case, is of independent theoretical interest. In the prior literature, it was not even known that the number of extreme points is finite.

4.3 Tracing the frontier

Ultimately, we will use basic pairs to design a pencil-sharpening algorithm for computing equilibrium payoffs. This algorithm will construct a sequence of pivot payoffs, each of which is generated by what is essentially an approximation of a basic pair, and the trajectory of these pivots will asymptotically converge to the frontier of $\mathbf{V}$. Before describing that algorithm in detail, it is useful to develop intuition by thinking about how one would traverse the frontier of $\mathbf{V}$ itself by pivoting between basic pairs, in an analogous fashion to the procedure we described in Section 3 for tracing the feasible payoff correspondence by pivoting between stationary strategies. Note that while the following discussion is informal, it will be rigorously justified by our formal results in Section 5, since we are essentially describing the behavior of our algorithm after it has converged to $\mathbf{V}$.

Suppose we are given an initial basic pair $(\mathbf{a}, \mathbf{r})$ that generates equilibrium payoffs $\mathbf{v}$ that are d-maximal (but not necessarily clockwise d-maximal). Let us suppose that we also know the sets of extreme binding continuation values $C(a)$ for all a. Even so, there is a large number of possible configurations of the basic pair, only some of which will generate maximal tuples in $\mathbf{V}$. We may therefore ask, as we did before, how we might use this information to identify other basic pairs that generate payoffs that are maximal for directions close to d. It turns out that we can do so using an appropriate generalization of the idea of "test directions" from Section 3. Consider, as before, the shallowest clockwise tangent d^{T} from $\mathbf{v}$ to $\mathbf{V}$ (under the assumption that $\mathbf{V} \neq \{\mathbf{v}\}$, so that this direction exists), and let $\mathbf{v}^{\mathrm{T}}$ denote the clockwise d^{T}-maximal equilibrium payoffs. Let us again focus on the state s^* in which the movement from $\mathbf{v}(s)$ to $\mathbf{v}^{\mathrm{T}}(s)$ is the largest. Since $\mathbf{v}^{\mathrm{T}}$ is clockwise d^{T}-maximal, Proposition 1 tells us that it is generated by a basic pair $\left(\mathbf{a}^{\mathrm{T}}, \mathbf{r}^{\mathrm{T}}\right)$.

Now suppose that $\mathbf{r}^{\mathrm{T}}(s^*) = \mathrm{NB}$, i.e., $\mathbf{v}^{\mathrm{T}}(s^*)$ is generated in the non-binding case. Thus,

$$\mathbf{v}^{\mathrm{T}}(s^*) = (1-\delta)\, g\left(\mathbf{a}^{\mathrm{T}}(s^*)\right) + \delta \sum_{s' \in S} \pi\left(s' | \mathbf{a}^{\mathrm{T}}(s^*)\right) \mathbf{v}^{\mathrm{T}}(s').$$

This situation is *exactly analogous* to that of the $\mathbf{v}$ and $\mathbf{v}^{\mathrm{T}}$ from Section 3, in which the next-clockwise payoffs were generated by using the same clockwise-maximal tuple as continuation values. In that setting, we saw that playing $\mathbf{a}^{\mathrm{T}}(s^*)$ for one period, followed by continuation payoffs $\mathbf{v}$, will generate a payoff that is between $\mathbf{v}(s^*)$ and $\mathbf{v}^{\mathrm{T}}(s^*)$. We can do the same in

the present setting: the *non-binding test direction*

$$d^{\mathrm{NB}}(a) = (1-\delta)\, g(a) + \delta \sum_{s' \in S} \pi(s'|a)\, \mathbf{v}(s') - \mathbf{v}(s)$$

is essentially what we called the test direction for feasible payoffs, and if we were to pick $s = s^*$ and $a = \mathbf{a}^{\mathrm{T}}(s^*)$, then it would point precisely in the direction d^{T}.

But what about when $\mathbf{r}^{\mathrm{T}}(s^*) \neq \mathrm{NB}$, so that $\mathbf{v}^{\mathrm{T}}(s^*)$ is generated with extreme binding continuation values? In this case, there need not be a non-binding direction that points towards $\mathbf{v}^{\mathrm{T}}$. We can, however, compute a small number of additional directions, one of which will surely point in the direction d^{T}. The *binding test directions* point from $\mathbf{v}(s)$ to the payoffs that can be generated by one of the (at most four) extreme binding continuation values $w \in C(a)$:

$$d^{\mathrm{B}}(a) = (1-\delta)\, g(a) + \delta w - \mathbf{v}(s).$$

If we pick $s = s^*$, $a = \mathbf{a}^{\mathrm{T}}(s^*)$, and if we consider the correct extreme binding continuation value, then the binding direction will point in same direction as d^{T}.

Thus, if we simply search over all of the binding and non-binding test directions, we will necessarily encounter the critical state, action, and possibly extreme binding continuation values that generate a test direction that points along the frontier. However, if we search over the test directions as we have just defined without further restrictions, we may even encounter test directions that point *above* d^{T} and out of $\mathbf{V}$ altogether. The reason is that the payoffs pointed to by $d^{\mathrm{NB}}(a)$ use $\mathbf{v}$ as continuation values, regardless of whether or not $\mathbf{v}$ is incentive compatible for a. It turns out that if we simply exclude those non-binding directions that are not incentive compatible, then the shallowest among the remaining (incentive compatible) non-binding directions and the binding directions will necessarily be parallel to d^{T}.[13]

In addition, one can use this shallowest test direction to identify a new basic pair that generates payoffs that move in the d^{T} direction relative to $\mathbf{v}$. Test directions are associated with continuation regimes in the obvious manner: $d^{\mathrm{NB}}(a)$ is associated with $r = \mathrm{NB}$ and $d^{\mathrm{B}}(a, w)$ is associated with $r = w$. We can therefore imagine making a substitution, as we did in Section 3, by simply changing the action and regime in the state that generates the shallowest direction to the (a, r) associated with that shallowest direction, to construct a new basic pair $(\mathbf{a}', \mathbf{r}')$. We can even compute new payoffs by iteratively applying the operator $\mu(\cdot|\mathbf{a}', \mathbf{r}')$ from equation (5) to the original payoff tuple $\mathbf{v}$. On the first application, payoffs will only move (in the direction d^{T}) in the state in which we made the substitution. On

[13] There is a subtlety here: even if $\mathbf{r}^{\mathrm{T}}(s^*) = \mathrm{NB}$, that is no guarantee that $\mathbf{v}$ is incentive compatible for $\mathbf{a}^{\mathrm{T}}(s^*)$. One can show, however, that in this situation there will be a binding test direction that points towards $\mathbf{v}^{\mathrm{T}}(s^*)$.

subsequent applications, payoffs will only move in for those states in which $\mathbf{r}'(s) = \text{NB}$. The linearity of μ in the continuation payoffs $\mathbf{w}$ implies that the generated payoffs move even further in the direction d^{T}, exactly as in the case of feasible payoffs.

There is, however, a new subtlety that arises with this Bellman procedure that did not appear in Section 3: it may be that over the course of iteratively applying μ, the iterates $\mu^k(\mathbf{v}; \mathbf{a}', \mathbf{r}')$ cease to be incentive compatible for some $\mathbf{a}'(s)$ for which $\mathbf{r}'(s) = \text{NB}$. If this happens, it means that $(\mathbf{a}', \mathbf{r}')$ is not incentive compatible per se. There is a simple modification, though, that will restore incentive compatibility. Note that if at one iteration the continuation values are incentive compatible, but at the next they are not, then the expected continuation value must "pass through" an incentive constraint. In addition, all of the payoffs generated by iteratively applying μ are d^{T}-maximal in $\mathbf{V}$. This means that the point at which the iterates cross the constraint is in fact an extreme binding continuation value. If we change $\mathbf{r}'(s)$ to be equal to this extreme binding continuation value, incentive compatibility will be restored, and as we continue to apply μ, payoffs will continue to move further in the d^{T} direction for any remaining non-binding states. These switches from non-binding to binding regimes can happen only finitely many times, and one can show that the limit of the sequence of payoff tuples $\mathbf{v}'$ must be generated by the appropriately modified basic pair $(\mathbf{a}', \mathbf{r}')$. Moreover, the limit payoffs $\mathbf{v}'$ necessarily lie between $\mathbf{v}$ and $\mathbf{v}^{\text{T}}$, and in many cases are equal to the latter. From this new basic pair and payoffs, one could again search over test directions, substitute in the shallowest action and regime, and compute a new basic pair that generates new basic payoffs. By repeating these steps over and over, the generated sequence of basic payoffs will march clockwise around the frontier of $\mathbf{V}$, and after finitely many repetitions will visit all of the clockwise maximal equilibrium payoffs.

The algorithm that we develop in the next section is essentially the generalization of this procedure to the case where we do not know the equilibrium threat point $\underline{\mathbf{v}}$ or the extreme binding equilibrium continuation values $C(a)$, but instead we replace these quantities with generous approximations, associated with some payoff correspondence that contains $\mathbf{V}$. The algorithm pivots between approximate basic pairs by generating test directions, and making the substitution associated with the shallowest test direction that is incentive compatible. The pivot payoffs move clockwise around $\mathbf{V}$, and as they do, we "shave of" parts of the approximation that are outside of the trajectory of the pivot. This shaving removes payoffs from the approximation that cannot be inductively generated. In the limit, only those payoffs survive which can be perpetually bootstrapped, i.e., those that can arise in equilibrium.

5 Calculating equilibrium payoffs

5.1 Test directions

Without further ado, we describe the general algorithm. We will begin by formally describing the appropriate generalizations of basic pairs and test directions to the case where we only have an approximation of $\mathbf{V}$. Suppose that $\mathbf{W}$ is a compact and convex payoff correspondence that contains $\mathbf{V}$. Also, suppose that $\mathbf{v} \in \mathbf{W}$ is a tuple of payoffs that are above $\mathbf{V}$ in some direction $\widehat{d}$. In other words, if we let

$$H(\mathbf{v}, d) = \left\{ \mathbf{w} | \mathbf{w}(s) \cdot \widehat{d} \leq \mathbf{v}(s) \cdot \widehat{d} \; \forall s \in S \right\}$$

denote the "half correspondence" of payoff tuples that are below the line $\mathbf{v} + xd$, then $\mathbf{V} \subseteq H(\mathbf{v}, d)$. The present question is, can we identify a direction in which the payoffs $\mathbf{v}$ can move that satisfies the following:

P1: (Non-zero) The direction is non-zero;

P2: (Containment) The direction does not point into the interior of $\mathbf{V}$;

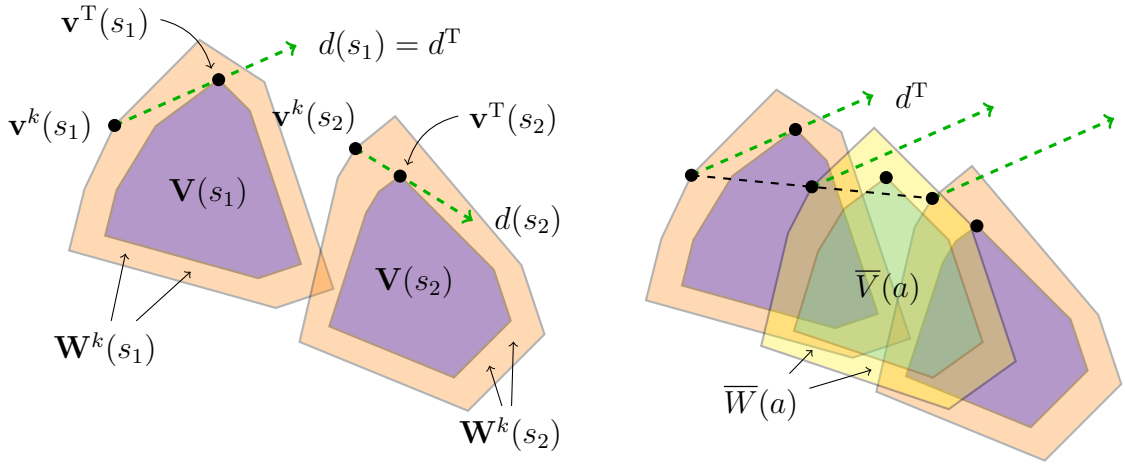
P3: (Feasible and incentive compatible) The direction points towards payoffs that can be generated from $\mathbf{W}$; and

P4: (Monotonicity) The direction points into $\mathbf{W}$ and below d.

If a direction satisfies all of **P1**–**P4**, then we will simply say it satisfies **P**. Properties **P1**–**P3** are natural desiderata given the discussion at the end of Section 4, and the reason for **P4** will be seen shortly, at which point we will give a formal definition. It turns out that we always can find a direction that satisfies **P**, by considering a series of *test directions* that generalize those of Section 3.

Let us establish some useful definitions. For each state, there is a clockwise tangent direction from $\mathbf{v}(s)$ to $\mathbf{V}(s)$, which we denote by $d(s)$. We take this direction to be zero if $\mathbf{V}(s) = \{\mathbf{v}(s)\}$, but if $\mathbf{V}(s)$ is not a singleton then this direction must be non-zero.[14] Let d^{T} denote the shallowest of all of these tangents across all states, which is attained in states in $S^{\mathrm{T}} \subseteq S$, and recall that $\widehat{d}^{\mathrm{T}}$ is the counter-clockwise normal to d^{T}. We write $\mathbf{v}^{\mathrm{T}}$ for the payoff tuple that is clockwise d^{T}-maximal in $\mathbf{V}$, and note that $\mathbf{v}^{\mathrm{T}}(s)$ is the tangent point from $\mathbf{v}(s)$ to $\mathbf{V}(s)$ for $s \in S^{\mathrm{T}}$. We depict these objects in Figure 6(a). Figure 6(b) depicts expected

[14] Note that when $\mathbf{v}(s) \in \mathbf{V}(s)$, the very existence of these tangent directions, and the fact that they are non-zero, relies upon the prior conclusion that each $\mathbf{V}(s)$ has finitely many extreme points (cf. Corollary 1).



(a) Tangent directions.

(b) Expected continuations.

Figure 6: The tangent directions for an example with two states $S = \{s_1, s_2\}$. $d^{\mathrm{T}} = d(s_1)$ is the shallowest tangent among all of the $d(s)$.

continuation values and provides useful context for the figures and arguments below. Thus, a sufficient condition for a direction to satisfy (i) is that it points above d^{T}.

When the action is a, the set of feasible expected continuation values is

$$\overline{W}(a) = \sum_{s' \in S} \pi(s'|a)\mathbf{W}(s').$$

Incentive compatibility is measured relative to the threat point $\underline{\mathbf{w}} = \underline{\mathbf{w}}(\mathbf{W})$. Thus,

$$h_i(a) = \max_{a_i'} \left[\frac{1-\delta}{\delta} \left(g_i(a_i', a_j) - g_i(a) \right) + \sum_{s' \in S} \pi(s'|a_i', a_j) \underline{\mathbf{w}}(s') \right]$$

and

$$IC(a) = \left\{ w \in \mathbb{R}^2 | w \geq h(a) \right\}$$

are respectively the minimum incentive compatible expected continuation value and the set of incentive compatible payoffs for action a. Finally, let

$$C(a) = \mathrm{ext}\left(\overline{W}(a) \cap \mathrm{bd} IC(a) \right) \tag{6}$$

denote the set of extreme feasible and binding expected continuation values. Finally, let

$$Gen(a) = (1-\delta) g(a) + \delta \left(\overline{W}(a) \cap IC(a) \right)$$

denote the set of payoffs that can be generated from a promising incentive compatible continuation values in $\mathbf{W}$, which is empty in the event that $\overline{W}(a) \cap IC(a)$ is empty.

The test directions fall into three categories. First, let

$$\widetilde{w}(a) = \sum_{s' \in S} \pi(s'|a)\mathbf{v}(s')$$

denote the expected pivot when actions a are played, and let

$$\widetilde{v}(a) = (1-\delta)g(a) + \delta\widetilde{w}(a)$$

denote the payoff that is generated by playing a today and going to continuation values $\mathbf{v}$ tomorrow. The *non-binding test direction* is

$$d^{\mathrm{NB}}(a) = \widetilde{v}(a) - \mathbf{v}(s).$$

The non-binding direction is associated with a regime $r = \mathrm{NB}$, and it is considered *incentive compatible* if $\widetilde{w}(a) \in IC(a)$.

The second type is the *binding test direction*, and is of the form

$$d^{\mathrm{B}}(a, w) = (1-\delta)g(a) + \delta w - \mathbf{v}(s),$$

for $w \in C(a)$. The binding test direction $d^{\mathrm{B}}(a, w)$ is associated with a regime $r = (\mathrm{B}, w)$, and it is always incentive compatible.

Together we refer to the non-binding and binding test directions as the *regular test directions*. These test directions are directly motivated by the structure of the basic pair, and they are the primary paths that will be considered by our algorithm. It turns out that we can always identify a regular test direction that satisfies **P1**–**P3** of our desiderata:

Lemma 4 (Regular test directions)**.** *Suppose that* $\mathbf{v} \in \mathbf{W}$*,* $\mathbf{V} \subseteq \mathbf{W} \cap H(\mathbf{v}, d)$*, and* $\mathbf{V} \neq \{\mathbf{v}\}$*. Then there exists a state* $s^* \in S$ *and actions* $a^* \in \mathbf{A}(s^*)$ *such that.*

(a) Either $d^{NB}(a^*) \cdot \widehat{d}^T > 0$ *or* $d^{NB}(a^*) = xd^T$ *for some* $x > 0$*. As a result,* $d^{NB}(a^*) \neq 0$*.*

(b) If $d^{NB}(a^*)$ *is not IC, then there exists a* $w^* \in C(a^*)$ *such that either* $d^B(a^*, w^*) \cdot \widehat{d}^T > 0$ *or* $d^B(a^*, w^*) = xd^T$ *for some* $x > 0$*. As a result,* $d^B(a^*, w^*) \neq 0$*.*

Proof of Lemma 4. We first identify the action a^*. Note that for states in S^{T}, we must be able to write

$$\mathbf{v}^{\mathrm{T}}(s) - \mathbf{v}(s) = \mathbf{x}(s)\, d^{\mathrm{T}}$$

for some $\mathbf{x}(s) > 0$. Let s^* denote any state in which $\mathbf{x}(s)$ attains its maximum, and let $\left(a^*, w^{\mathrm{T}}\right)$ denote the action pair and continuation value that generate $\mathbf{v}^{\mathrm{T}}(s^*)$:

$$\mathbf{v}^{\mathrm{T}}(s^*) = (1-\delta)\, g(a^*) + \delta w^{\mathrm{T}}.$$

In addition, define

$$\begin{aligned} \widetilde{w} &= \sum_{s \in S} \pi(s|a^*)\, \mathbf{v}(s); \\ \widetilde{v} &= (1-\delta)\, g(a^*) + \delta \widetilde{w}; \\ d^{\mathrm{NB}} &= \widetilde{v} - \mathbf{v}(s^*), \end{aligned}$$

which are, respectively, the expected pivot when a^* is played, the payoff that is generated by playing a^* for one period and using $\mathbf{v}$ as continuation values, and the non-binding direction associated with a^*. Also, let w^* denote the clockwise d^{T}-maximal element of $C(a^*)$, and define

$$\begin{aligned} v^* &= (1-\delta)\, g(a^*) + \delta w^*; \\ d^{\mathrm{B}} &= v^* - \mathbf{v}(s^*), \end{aligned}$$

which are, respectively, the payoff generated by (a^*, w^*) and the "best" binding direction associated with a^*, in the sense of smallest clockwise angle relative to d^{T}.

We now prove (a). Since d^{T} is the shallowest tangent, we must have

$$\mathbf{v}(s) \cdot \widehat{d}^{\mathrm{T}} \geq \mathbf{v}'(s) \cdot \widehat{d}^{\mathrm{T}} \tag{7}$$

for all $\mathbf{v}' \in \mathbf{V}$. As a result,

$$\widetilde{w} \cdot \widehat{d}^{\mathrm{T}} \geq w \cdot \widehat{d}^{\mathrm{T}} \tag{8}$$

for all $w \in \overline{V}(a^*)$. Since $w^{\mathrm{T}} \in \overline{V}(a^*)$, we conclude that

$$\begin{aligned} \widetilde{v} \cdot \widehat{d}^{\mathrm{T}} &= (1-\delta)\, g(a^*) \cdot \widehat{d}^{\mathrm{T}} + \delta \widetilde{w} \cdot \widehat{d}^{\mathrm{T}} \\ &\geq (1-\delta)\, g(a^*) \cdot \widehat{d}^{\mathrm{T}} + \delta w^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{T}} \\ &= \mathbf{v}^{\mathrm{T}}(s^*) \cdot \widehat{d}^{\mathrm{T}}, \end{aligned}$$

and hence

$$d^{\mathrm{NB}} \cdot \widehat{d}^{\mathrm{T}} \geq 0. \tag{9}$$

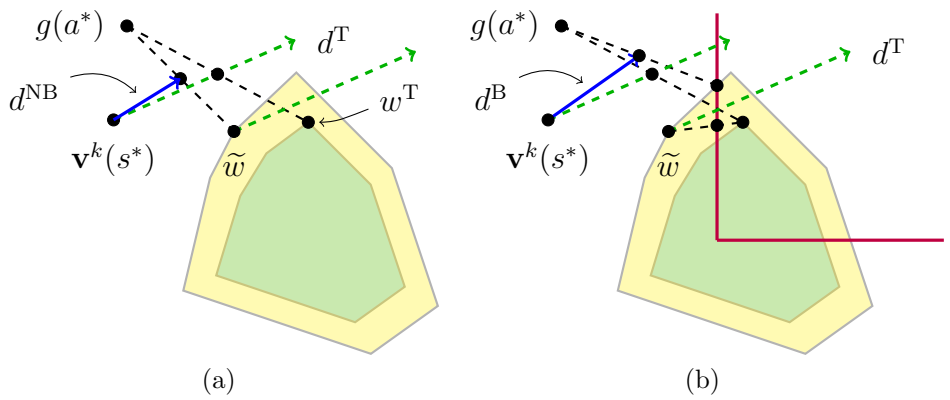


Figure 7: (a) The non-binding direction is always shallower than d^{T}. (b) The non-binding direction fails **P3**, and the shallowest binding direction is shallower than d^{T}.

The geometry of the non-binding test direction is depicted in Figure 7(a).

If (9) is a strict inequality, then we are done. Otherwise, then the fact that $d^{\mathrm{NB}} \cdot \widehat{d}^{\mathrm{T}} = 0$ implies that $\widetilde{w} \cdot \widehat{d}^{\mathrm{T}} = w^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{T}}$. Since $w^{\mathrm{T}} \in \overline{W}(a^*)$, it must be possible to write

$$w^{\mathrm{T}} = \sum_{s' \in S} \pi(s'|a^*)\, \mathbf{w}(s')$$

for some $\mathbf{w} \in \mathbf{V}$, and moreover, due to (7) and the definition of $\widetilde{w}$, it must be that

$$\mathbf{w}(s) - \mathbf{v}(s) = \mathbf{y}(s)\, d^{\mathrm{T}} \tag{10}$$

for all s in $S_+ = \{s|\pi(s|a^*) > 0\}$. Finally, since $\mathbf{v}^{\mathrm{T}}$ is clockwise d^{T}-maximal in $\mathbf{V}$, it must be that

$$\mathbf{y}(s) \leq \mathbf{x}(s) \leq \mathbf{x}(s^*) \tag{11}$$

for all $s \in S_+$. This implies that

$$\begin{aligned} d^{\mathrm{NB}} &= \widetilde{v} - \mathbf{v}(s^*) \\ &= \mathbf{v}^{\mathrm{T}}(s^*) - \mathbf{v}(s^*) - \delta \sum_{s \in S} \pi(s|a^*)\left(\mathbf{w}(s) - \mathbf{v}(s)\right) \\ &= \left(\mathbf{x}(s^*) - \delta \sum_{s \in S_+} \pi(s|a^*)\, \mathbf{y}(s)\right) d^{\mathrm{T}}, \end{aligned}$$

where the coefficient on d^{T} must be strictly positive because $\delta < 1$ and equation (11). This proves the first part of the Lemma.

We now prove (b). We have shown that d^{NB} is non-zero. Now suppose that it is not incentive compatible. From the definitions of d^{B} and $\mathbf{x}$, we can write

$$d^{\mathrm{B}} = \mathbf{x}\left(s^*\right) d^{\mathrm{T}} + \delta\left(w^* - w^{\mathrm{T}}\right).$$

We will argue that $w^* \cdot \widehat{d}^{\mathrm{T}} \geq w^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{T}}$, which will imply the result. Since $\widetilde{w}$ and w^{T} are both in $\overline{W}\left(a^*\right)$, so is every convex combination $w\left(\alpha\right) = \alpha\widetilde{w} + \left(1-\alpha\right) w^{\mathrm{T}}$. Also, since $w^{\mathrm{T}} \in IC\left(a^*\right)$ (it is in fact incentive compatible for the even more stringent threats $\underline{\mathbf{v}}$) and $\widetilde{w} \notin IC\left(a^*\right)$, there must exist some α^* such that $w\left(\alpha^*\right) \in \mathrm{bd}IC\left(a^*\right) \cap \overline{W}\left(a^*\right)$. This implies that w^*, the clockwise d^{T}-maximal element of $\mathrm{bd}IC\left(a^*\right) \cap \overline{W}\left(a^*\right)$, satisfies

$$\begin{aligned} w^* \cdot \widehat{d}^{\mathrm{T}} &\geq w\left(\alpha^*\right) \cdot \widehat{d}^{\mathrm{T}} \\ &\geq \left(1-\alpha^*\right) \widetilde{w} \cdot \widehat{d}^{\mathrm{T}} + \alpha^* w^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{T}} \\ &\geq w^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{T}} \end{aligned}$$

as desired. This geometry is depicted in Figure 7(b).

If this inequality is strict, we are done. Otherwise, it must be that

$$w^* \cdot \widehat{d}^{\mathrm{T}} = \widetilde{w} \cdot \widehat{d}^{\mathrm{T}} = w^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{T}},$$

and indeed

$$w^* = \left(1-\alpha^*\right) \widetilde{w} + \alpha^* w^{\mathrm{T}}.$$

Moreover, in this case we know that $d^{\mathrm{NB}} = x d^{\mathrm{T}}$ for some $x > 0$. This implies that

$$\begin{aligned} d^{\mathrm{B}} &= \left(1-\alpha^*\right) \widetilde{v} + \alpha^* \mathbf{v}^{\mathrm{T}}\left(s^*\right) - \mathbf{v}\left(s^*\right) \\ &= \left(\left(1-\alpha^*\right) x + \alpha^* \mathbf{x}\left(s^*\right)\right) d^{\mathrm{T}} \end{aligned}$$

as claimed. □

Thus, it is always possible to find actions a^* that generate a regular test direction, which we will call $d^{*,\mathrm{R}}$, that satisfies **P1**–**P3**. In other words, $d^{*,\mathrm{R}}$ is shallower than the shallowest tangent, and therefore will not cut into the interior of $\mathbf{V}$, and it also points towards payoffs that can be generated by promising continuation values in $\mathbf{W}$. Specifically, we define $d^{*,\mathrm{R}}$ to be $d^{\mathrm{NB}}\left(a^*\right)$ if that direction is incentive compatible, and otherwise $d^{*,\mathrm{R}}$ is equal to $d^{\mathrm{B}}\left(a^*, w^*\right)$.

Notice, however, that this conclusion required us to assume that $\mathbf{v}$ was itself contained in $\mathbf{W}$. The critical step in the argument was when we concluded that there exists a binding

direction that is shallower than d^{T} in the event that $d^{\mathrm{NB}}(a^*)$ is not incentive compatible. Thus, when we ultimately adapt these test directions to construct an algorithm for computing $\mathbf{V}$, we will require that the pivot always stay inside the approximation $\mathbf{W}$. As a result, we want the direction selected by the algorithm to be *monotonic*, in the sense that $\mathbf{v}(s) + xd^* \in \mathbf{W}$ for some $x > 0$.

Unfortunately, we have no guarantee that the direction $d^{*,\mathrm{R}}$ identified by Lemma 4 will satisfy **P4**, and it is a generic possibility (for an arbitrarily chosen $\mathbf{W}$and $\mathbf{v}$) that this direction could point out of $\mathbf{W}$. This may be surprising, since the APS operator is monotonic. However, the non-binding and binding directions point to a relatively small subset of the payoffs that can be generated from $\mathbf{W}$, specifically those which are candidates to be extremal payoffs according to the characterization of Section 4, and these payoffs can move in complicated ways as $\mathbf{W}$ changes.[15]

Thus, in the event that the critical regular test direction is non-monotonic, we will replace it with another test direction which satisfies all of $\mathbf{P}$. For a given action pair a in state s, let v^{CW} denote the next clockwise maximal payoffs from $\mathbf{v}(s)$ along the frontier of $\mathbf{W}(s) \cap H(\mathbf{v}(s), d)$, and let

$$d^{\mathrm{CW}} = v^{\mathrm{CW}} - \mathbf{v}(s),$$

i.e., d^{CW} is the clockwise tangent from $\mathbf{v}(s)$ to $\mathbf{W} \cap H(\mathbf{v}, d)$. This vector is proportional to d except when $\mathbf{v}(s)$ is an extreme point of $\mathbf{W} \cap H(\mathbf{v}, d)$. Further, let

$$x^{\mathrm{C}} = \max\left\{x | \mathbf{v}(s) + xd^{\mathrm{CW}} \in Gen(a)\right\},$$

when the set over which we are maximizing is non-empty. Finally, when $x^{\mathrm{C}} > 0$, we define the *continuation test direction* to be

$$d^{\mathrm{C}}(a) = \min\left\{x^{\mathrm{C}}, 1\right\} d^{\mathrm{CW}}.$$

We associate the continuation direction with a regime $r = (\mathrm{C}, v)$, where $v = \mathbf{v}(s) + d^{\mathrm{C}}(a)$.

Lemma 5 (Continuation direction)**.** *Let a^* and $d^{*,R}$ be the action profile and direction identified in the proof of Lemma 4, and suppose that $d^{*,R}$ is non-monotonic. Then $x^C > 0$, so that the continuation direction $d^C(a^*)$ is monotonic and satisfies **P**.*

Proof of Lemma 5. Let us argue that x^{C} is strictly positive. Since $\mathbf{V}$ is contained in $\mathbf{W}$, it must be that d^{T} is below d^{CW}, in the sense that $d^{\mathrm{CW}} \cdot \widehat{d}^{\mathrm{T}} > 0$ or $d^{\mathrm{T}} = yd^{\mathrm{CW}}$ for some $y > 0$.

[15]We note that these non-monotonicities also occurred in Abreu and Sannikov (2014). We further note that such non-monotonicities cannot happen when $\mathbf{W} = \mathbf{V}$, i.e., when we are at the fixed point, for the existence of a non-monotonic regular test direction would imply that there are equilibrium payoffs that can be generated from $\mathbf{V}$ that are outside of $\mathbf{V}$.

In the latter case, x^C must be positive because $\mathbf{v}^{\mathrm{T}}(s^*)$ is itself an element of $Gen(a^*)$ and

$$\begin{aligned}\mathbf{v}^{\mathrm{T}}(s^*) &= \mathbf{v}(s^*) + \mathbf{x}(s^*) d^{\mathrm{T}} \\ &= \mathbf{v}(s^*) + y\mathbf{x}(s^*) d^{\mathrm{CW}}.\end{aligned}$$

So suppose that $d^{\mathrm{CW}} \cdot \widehat{d}^{\mathrm{T}} > 0$, so that the tangent direction points strictly below d^{CW}. In that case, we can find non-zero weights (x, y) such that

$$d^{*,\mathrm{R}} = xd^{\mathrm{CW}} + yd^{\mathrm{T}},$$

and since $d^{*,\mathrm{R}} \cdot \widehat{d}^{\mathrm{T}} > 0$ (if this were zero $d^{*,\mathrm{R}}$ would be monotonic, and thus satisfy **P**), we conclude that $xd^{\mathrm{CW}} \cdot \widehat{d}^{\mathrm{T}} > 0$, which is only possible if $x > 0$. If $y \geq 0$, then $d^{*,\mathrm{R}}$ must point into $\mathbf{W}(s)$, for we can find a z sufficiently small so that

$$\mathbf{v}(s) + zxd^{\mathrm{CW}} \in \mathbf{W}(s)$$

and such that $zy < \mathbf{x}(s^*)$, so that $\mathbf{v}(s^*) + zd^*$ is a convex combination of a point on the clockwise edge emanating from $\mathbf{v}(s^*)$ and a point between $\mathbf{v}(s^*)$ and $\mathbf{v}^{\mathrm{T}}(s^*)$. Thus, it must be that $y < 0$, so that

$$\begin{aligned}d^{*,\mathrm{R}} \cdot \widehat{d}^{\mathrm{CW}} &= yd^{\mathrm{T}} \cdot \widehat{d}^{\mathrm{CW}} \\ &= -y\widehat{d}^{\mathrm{T}} \cdot d^{\mathrm{CW}} > 0.\end{aligned}$$

Let us write

$$v = \mathbf{v}(s^*) + d^*$$

for the payoffs pointed to by $d^{*,\mathrm{R}}$. Since $d^{*,\mathrm{R}}$ points above d^{CW} and d^{T} points below d^{CW}, it must be that

$$v \cdot \widehat{d}^{\mathrm{CW}} > \mathbf{v}(s^*) \cdot \widehat{d}^{\mathrm{CW}} \geq \mathbf{v}^{\mathrm{T}}(s^*) \cdot \widehat{d}^{\mathrm{CW}}.$$

Note that both v and $\mathbf{v}^{\mathrm{T}}(s^*)$ can be generated by a^*, so that for every $\alpha \in [0, 1)$, the payoff $v(\alpha) = \alpha v + (1 - \alpha)\mathbf{v}^{\mathrm{T}}$ is contained in $Gen(a^*)$. Moreover, we can find a critical α^* such that $v(\alpha^*)$ lies on exactly the same d^{CW} plane as $\mathbf{v}(s^*)$. As a result,

$$\begin{aligned}v(\alpha^*) - \mathbf{v}(s^*) &= \alpha^* d^{*,\mathrm{R}} + (1 - \alpha^*) d^{\mathrm{T}} \\ &= \alpha^* xd^{\mathrm{CW}},\end{aligned}$$

so that x^{C} in the definition of the continuation direction is strictly positive. □

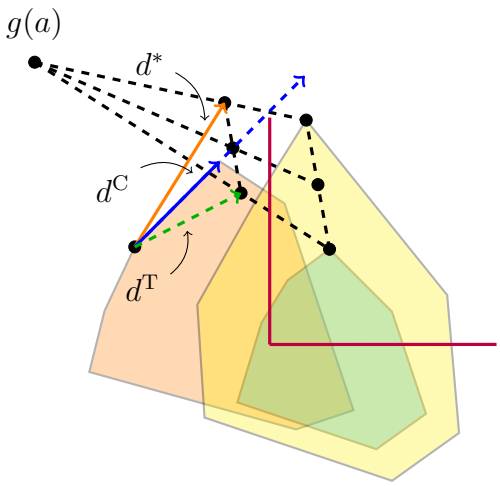


Figure 8: The continuation direction.

We conclude from Lemmas 4 and 5 that there is an action a^* that generates a direction d^* that satisfies **P**. In particular, the direction is non-zero, it is shallower than the shallowest tangent, it points to payoffs that can be generated from **W**, and it is monotonic. The direction d^* is equal to $d^{*,\mathrm{R}}$ in the event that $d^{*,\mathrm{R}}$ is monotonic, and otherwise it is equal to $d^{\mathrm{C}}(a^*)$.

5.2 Finding a new approximate basic pair

Ultimately, we will use the test directions to iteratively identify approximate basic pairs that are generous estimates of the shape of **V**. Specifically, we redefine a basic pair $(\mathbf{a}, \mathbf{r})$ to be the analogous object from Section 4, except where feasibility and incentive compatibility are defined relative to **W** rather than **V**, and we also allow the possibility that regimes can take on the additional values (C, v) when the present pivot is the result of a continuation direction. The generalized basic pair induces unique payoffs **v** according to

$$\mathbf{v}(s) = \begin{cases} (1-\delta)\, g(\mathbf{a}(s)) + \delta \sum_{s' \in S} \pi(s'|\mathbf{a}(s))\, \mathbf{v}(s') & \text{if } \mathbf{r}(s) = \mathrm{NB}; \\ (1-\delta)\, g(\mathbf{a}(s)) + \delta w & \text{if } \mathbf{r}(s) = (\mathrm{B}, w); \\ v & \text{if } \mathbf{r}(s) = (\mathrm{C}, v), \end{cases} \tag{12}$$

and the pair $(\mathbf{a}, \mathbf{r})$ is *incentive compatible* if, for all $s \in S$ in which $\mathbf{r}(s) = \mathrm{NB}$,

$$\sum_{s' \in S} \pi(s'|\mathbf{a}(s))\, \mathbf{v}(s') \in IC(\mathbf{a}(s)), \tag{13}$$

where again $IC(a)$ is defined relative to the approximate threats $\underline{\mathbf{w}}(\mathbf{W})$. In this case, we say that $(\mathbf{a}, \mathbf{r})$ *generates* $\mathbf{v}$ *from* $\mathbf{W}$, or that it just generates $\mathbf{v}$, when $\mathbf{W}$ is clear from context.

Now suppose that the pivot $\mathbf{v}$ from the previous subsection is generated by a basic pair $(\mathbf{a}, \mathbf{r})$. From the previous section, we know that there is a test direction d^*, generated by an action and regime (a^*, r^*)[16] in state s^*, that is shallower than the shallowest tangent d^{T}, points to payoffs that can be generated from $\mathbf{W}$, and is monotonic. We can use this information to find a new basic pair $(\mathbf{a}', \mathbf{r}')$ that generates payoffs

$$\mathbf{v}' = \mathbf{v} + \mathbf{x}d^*$$

for some non-negative scalars $\mathbf{x}(s)$ that are not all zero.

First, set

$$\mathbf{a}'(s) = \begin{cases} a^* & \text{if } s = s^*; \\ \mathbf{a}(s) & \text{otherwise.} \end{cases}$$

The new regime $\mathbf{r}'$ and movements $\mathbf{x}$ will be determined as the limits of sequences $\left\{\mathbf{r}^k\right\}_{k=0}^{\infty}$ and $\left\{\mathbf{x}^k\right\}_{k=0}^{\infty}$, where

$$\mathbf{r}^0(s) = \begin{cases} r^* & \text{if } s = s^*; \\ \mathbf{r}(s) & \text{otherwise,} \end{cases}$$

and

$$\mathbf{x}^0 = \begin{cases} 1 & \text{if } s = s^*; \\ 0 & \text{otherwise.} \end{cases}$$

We also let $\mathbf{x}^{-1}(s) = 0$ for all s.

To motivate the iteration, suppose we tried to generate payoffs from $(\mathbf{a}', \mathbf{r}^0)$. Verily, these payoffs will move in the direction d^* relative to $\mathbf{v}$. We have no guarantee, however, that the resulting tuple will be incentive compatible for states in which $\mathbf{r}^0(s) = \text{NB}$. It might be that the induced payoffs move *too* far in the direction d^*, so that they move outside of the incentive compatible region for the relevant states. In addition, the induced payoffs could move so far in the direction d^* that they move outside of $\mathbf{W}$ altogether. The iterative procedure gradually moves payoffs in the direction d^* for non-binding states to identify whether or not

[16] While we continue to use d^* to denote the shallow direction and a^* to denote the actions that generate it, we note that for the purposes of the following discussion and Proposition 2, it is *not* necessary that a^* and d^* be the particular actions and direction identified in Lemmas 4 and 5. Indeed, d^* can be any direction that satisfies conditions (i-iii) from the beginning of Section 5. This is essential to the operation of our algorithm, which may in general identify directions that satisfy (i-iii) and are shallower than the d^* of Lemmas 4 and 5.

constraints would be violated, and if so, change the regimes in the relevant states so that payoffs stop at the appropriate incentive and monotonicity constraints.

We now define the iterative procedure. At iteration $k > 0$, for all states in which $\mathbf{r}^k(s) \neq$ NB, we simply set

$$\left(\mathbf{r}^k(s), \mathbf{x}^k(s)\right) = \left(\mathbf{r}^{k-1}(s), \mathbf{x}^{k-1}(s)\right).$$

Otherwise, we compute

$$\overline{x} = \mathbf{x}^{k-1}(s) + \delta \sum_{s' \in S} \pi\left(s' | \mathbf{a}(s)\right) \left(\mathbf{x}^{k-1}(s') - \mathbf{x}^{k-2}(s')\right),$$

and let

$$x = \max\left\{y \leq \overline{x} | \mathbf{v}(s) + yd^* \in Gen\left(\mathbf{a}(s)\right) \cap \mathbf{W}(s)\right\}, \tag{14}$$

which is necessarily non-negative.

If $x = \overline{x}$, then we simply set $\mathbf{x}^k(s) = \overline{x}$. Otherwise, we set

$$\left(\mathbf{r}^k(s), \mathbf{x}^k(s)\right) = \left(\left(\mathrm{B}, w\right), x\right)$$

where

$$w = \frac{1}{\delta}\left(\mathbf{v}(s) + xd^* - (1-\delta)\, g\left(\mathbf{a}(s)\right)\right)$$

is the continuation value that generates $\mathbf{v}(s) + xd^*$.

This completes the specification. Let us define

$$\mathbf{r}' = \lim_{k \to \infty} \mathbf{r}^k, \quad \mathbf{x}' = \lim_{k \to \infty} \mathbf{x}^k.$$

The following proposition characterizes the procedure.

Proposition 2 (New basic pair)**.** *The basic pair* $(\mathbf{a}', \mathbf{r}')$ *in the preceding discussion is well defined. Moreover,* $(\mathbf{a}', \mathbf{r}')$ *generates basic equilibrium payoffs*

$$\mathbf{v}'(s) = \mathbf{v}(s) + \mathbf{x}(s)\, d^*.$$

Proof of Proposition 2. Let us first argue that our procedure is well defined. For any state such that $\mathbf{r}^0 =$ NB, the payoffs

$$\mathbf{v}(s) + \mathbf{x}^0(s)\, d^* \in Gen\left(\mathbf{a}'(s)\right).$$

If $s = s^*$, then $r^* =$ NB, so the non-binding test direction must have been IC. On the other hand, if $s \neq s^*$, then it must have been that $\mathbf{r}(s) =$ NB as well, and this conclusion

follows from the hypothesis that $\mathbf{v}(s)$ are generated by $(\mathbf{a}, \mathbf{r})$. Now, suppose inductively that $\mathbf{x}^{k-1} \geq \mathbf{x}^{k-2}$ and that

$$\mathbf{v}(s) + \mathbf{x}^{k-1}(s) d^* \in Gen(\mathbf{a}'(s)) \cap \mathbf{W}(s),$$

the base case having just been established for $k = 1$. Then it must be that x in equation (14) is well-defined and non-negative.

Finally, let us argue that the regime and movement sequences converge, and that $(\mathbf{a}', \mathbf{r}')$ generates $\mathbf{v} + \mathbf{x}' d^*$. While $\mathbf{r}^k$ is not changing, our procedure is essentially iteratively applying the Bellman operator of equation (5) which, as we have previously observed, is a contraction of modulus δ. Thus, the iterates $\mathbf{x}^k$ converge at a geometric rate to the unique fixed point $\mathbf{x}'$ which satisfies

$$\mathbf{x}'(s) - \mathbf{x}^0(s) = \delta \sum_{s' \in S} \pi(s'|\mathbf{a}'(s)) \mathbf{x}'(s').$$

Thus,

$$\mathbf{v}'(s) = \mathbf{v}(s) + \mathbf{x}^0(s) d^* + \delta \sum_{s' \in S} \pi(s'|\mathbf{a}'(s)) \mathbf{x}'(s') d^*.$$

Note that if $\mathbf{r}'(s) = \text{NB}$,

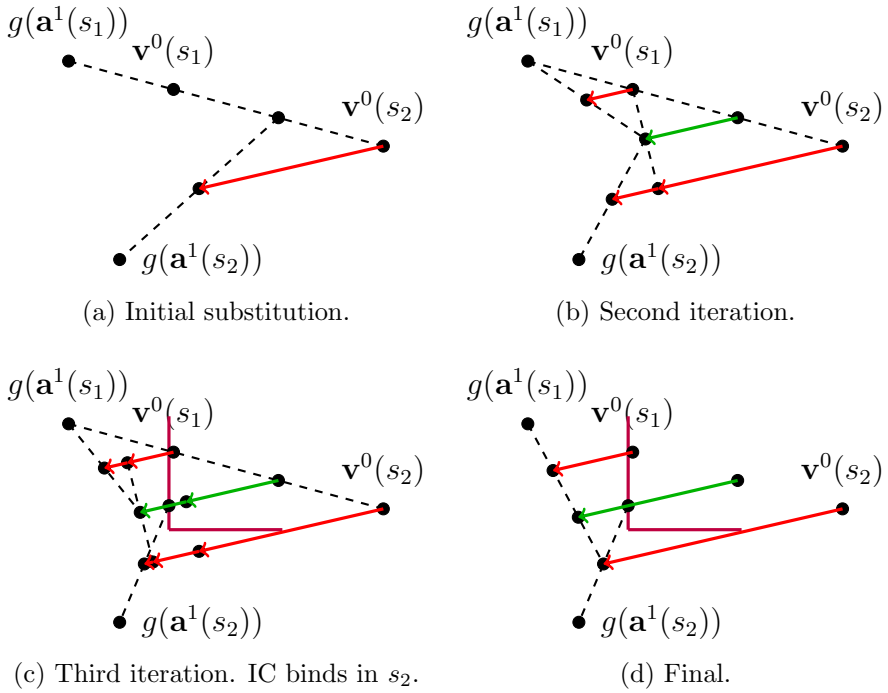
$$\mathbf{v}(s) + \mathbf{x}^0(s) d^* = (1-\delta) g(\mathbf{a}'(s)) + \delta \sum_{s' \in S} \pi(s'|\mathbf{a}'(s)) \mathbf{v}(s'),$$

so

$$\mathbf{v}'(s) = (1-\delta) g(\mathbf{a}(s)) + \delta \sum_{s' \in S} \pi(s'|\mathbf{a}'(s)) \mathbf{v}'(s').$$

Now suppose that $\left(\mathbf{a}', \mathbf{r}^k\right)$ generates payoffs that are not incentive compatible or are not contained in $\mathbf{W}$. In that case, after finitely many iterations, an incentive or a monotonicity constraint will be violated by $\mathbf{v} + \mathbf{x}^k d^*$, and $\mathbf{r}^k$ will be changed to a binding regime. Since there are only finitely many states, there can be only finitely many switches from non-binding to binding regimes, and $\mathbf{r}^k$ must converge after finitely many iterations. At this point, $\mathbf{x}^k$ converges to the fixed point at which incentive and monotonicity constraints are satisfied. □

This Bellman procedure is depicted graphically in Figure 9. In this example, $\mathbf{r}(s_1)$ is non-binding at every iteration, whereas $\mathbf{r}(s_2)$ starts out non-binding but eventually transitions to binding. In Figure 9(a), the initial substitution is made that moves payoffs in state s_2 in a south-westerly direction. Figure 9(b) shows the second iteration, in which the movement in state s_2 is propagated through to state s_1. Through the second iteration, the expected pivots



(a) Initial substitution.

(b) Second iteration.

(c) Third iteration. IC binds in s_2.

(d) Final.

Figure 9: The Bellman procedure.

are incentive compatible in both states. At iteration three, however, the incentive constraint in state s_2 would be violated by using the pivot as continuation values. As a result, we fix the payoff in s_2 at the binding constraint, but move the continuation payoff for state s_1 all the way to the expected pivot. This is depicted in Figure 9(c). In Figure 9(d), we see the final configuration of the new pivot.

Note that we have described this procedure as if there are infinitely many iterations. In practice, we have implemented this procedure by inverting the system of equations (12) for each $\left(\mathbf{a}^1, \mathbf{r}^{1,k}\right)$ and checking if the resulting payoffs satisfy the incentive constraint (13). If not, we iterate as above until a constraint is violated, update the regimes, and then invert again. After finitely many rounds, the payoffs obtained by inverting (12) must be incentive compatible.

5.3 The algorithm

The test directions and the updating procedure are the building blocks from which we will construct an algorithm for computing $\mathbf{V}$. The algorithm proceeds over a series of iterations, over the course of which we will generate a sequence of *pivot* payoff tuples $\mathbf{v}^k$ and accompanying action and regime tuples $\left(\mathbf{a}^k, \mathbf{r}^k\right)$. We will also keep track of a *current direction* d^k

that satisfies

$$\mathbf{V} \subseteq H\left(\mathbf{v}^k, d^k\right).$$

This means that the equilibrium payoff correspondence is always below $\mathbf{v}^k$ in levels with slope d^k. In addition, we will maintain a compact and convex payoff correspondence $\mathbf{W}^k$, which contains the equilibrium payoff correspondence and serves as our approximation of the payoffs that can be promised as binding

continuation values on the equilibrium path and as punishment continuation values after deviations. This $\mathbf{W}^k$ will in fact be the set of payoffs that have been circumscribed by the trajectory of the pivot $\mathbf{v}^k$ thus far, i.e.,

$$\mathbf{W}^k = \mathbf{W}^0 \cap \left(\cap_{l=0}^{k} H\left(\mathbf{v}^l, d^l\right)\right).$$

The algorithm can be initialized with any $\mathbf{v}^0$, d^0, and $\mathbf{W}^0$ that satisfy these conditions. The initial $\mathbf{a}^0$ and $\mathbf{r}^0$ can be arbitrary as long as $\mathbf{r}^0(s) \neq \text{NB}$ for all s.

At each iteration, we will search over all test directions according to a procedure that we will describe presently. The test direction which generates the smallest clockwise angle relative to d^k will be deemed *shallowest* and will become the new current direction d^{k+1}. We then substitute the action and regime that generated the best direction into the system (3) using the procedure described in Section 5.2, and advance the pivot to

$$\mathbf{v}^{k+1} = \mathbf{v}^k + \mathbf{x}d^{k+1}$$

where $\mathbf{x}$ is a tuple of non-negative scalars. Lemmas 4 and 5 will imply that d^{k+1} satisfies

$$\mathbf{V} \subseteq H\left(\mathbf{v}^k, d^{k+1}\right) = H\left(\mathbf{v}^{k+1}, d^{k+1}\right),$$

so that our approximations will continue to contain all of the equilibrium payoffs.

The algorithm proceeds over a sequence of such iterations, through which the pivot tuple moves clockwise, revolving around and around the equilibrium payoff sets. Our convention will be that the new revolution begins when the d^k passes due north $d^{\text{N}} = (0, 1)$, i.e., when d^{k-1} points somewhere to the west of due north, and d^k points somewhere to the east. The index of the iteration can therefore be decomposed as $k = r : c$, where r is the number of revolutions and c is the number of cuts, or steps, within the revolution. The current

revolution and cut are denoted by $r(k)$ and $c(k)$, respectively. With slight abuse of notation, we will write $k+1 = r+1 : 0$ if $k+1$ starts a new revolution and $k+1 = r : c+1$ otherwise.

Within iteration k, we search over the directions described in Section 5.1, where the current pivot and direction are $\mathbf{v}^k$ and d^k, and the feasible payoff correspondence is $\mathbf{W}^{r(k):0}$, i.e., the approximation at the *beginning* of the current iteration. We shall see that keeping the feasible payoff correspondence constant within a revolution simplifies our subsequent anti-stalling arguments. For each action pair a, we look for the direction that would be identified by Lemmas 4 and 5, under the hypothesis that this action pair is in fact the a^* that generates the shallowest tangent with the largest movement. We denote this candidate direction by $d^*(a)$. If we encounter conditions that imply that a *cannot* be a^*, then we simply set $d^*(a)$ to be null, i.e., $d^*(a) = \emptyset$. The algorithm then selects as the next direction the shallowest direction from among all of the non-null candidates:

$$D^* = \{d^*(a) \,|\, a \in A, d^*(a) \neq \emptyset\}.$$

The exact search procedure is portrayed as a tree in Figure 10. We know from Lemma 4 that the critical actions a^* *must* generate a non-binding direction that is above d^{T}. Thus, the first thing our search procedure does for each a is examine its associated non-binding test direction $d^{\mathrm{NB}}(a)$. Note that d^{T} must point into $\mathbf{W}^k$, so it must therefore be below both d^k and d^{CW}. Hence, if a is the critical action pair and if $d^{\mathrm{NB}}(a)$ is below d^{CW}, then $d^{\mathrm{NB}}(a)$ must be below d^{CW} and above d^{T} and therefore point into $\mathbf{W}^k$. If this is not the case, then we can simply skip this action and set $d^*(a) = \emptyset$. We can also rule out a being a^* if $d^{\mathrm{NB}} = 0$.

If $d^{\mathrm{NB}}(a)$ is non-zero and if it might be above d^{T}, then we check whether it is incentive compatible or not (cf. Lemma 4(a)). If the non-binding direction satisfies **P**, then it is a candidate to be the critical direction, and we set $d^*(a) = d^{\mathrm{NB}}(a)$. Otherwise, if it fails **P4**, then we can check for the continuation direction, as in Lemma 5. If no continuation direction exists, because there are no payoffs that can be generated in the direction d^{CW}, then it is impossible that this action pair is a^*. If $d^{\mathrm{C}}(a)$ does exist, then we take it as the candidate for the critical direction from a^*.

Finally, if we find that $d^{\mathrm{NB}}(a)$ is not incentive compatible, so that **P3** fails, then Lemma 4 tells us (again under the hypothesis that the actions under consideration are a^*) that there is a non-zero binding direction that is shallower than the shallowest tangent. We therefore look for the shallowest binding direction generated by a, and if this direction $d^{\mathrm{B}}(a, w)$ is monotonic, then we take it as the candidate to be the critical direction with $d^*(a) = d^{\mathrm{B}}(a, w)$. If it is non-monotonic, then again we invoke Lemma 5 and look for a continuation direction. If one exists, then we set it to be the candidate associated with a.

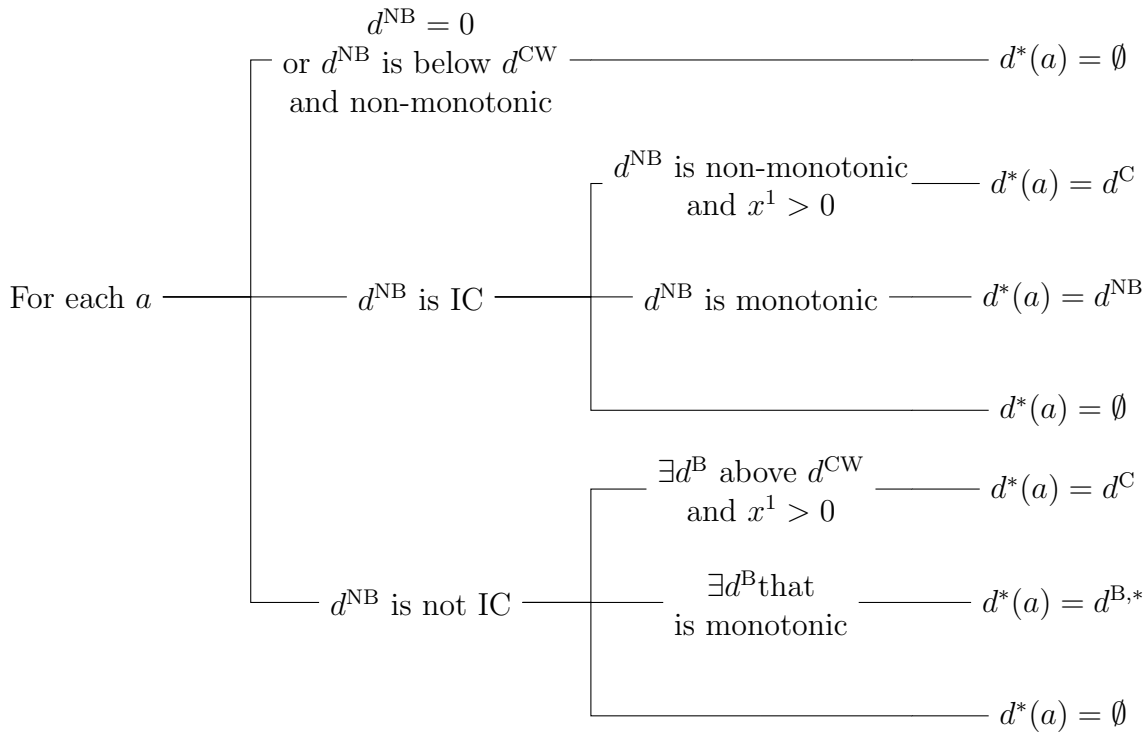


Figure 10: The search for the next direction.

Otherwise, a cannot be a^*, and we move on to the next action pair. This completes the specification of our search procedure.

5.4 Characterization of the algorithm

We now prove that the sequence of approximations $\mathbf{W}^k$ converge to the equilibrium payoff correspondence $\mathbf{V}$. This will be verified in three steps.

5.4.1 Containment

First, we will show that as long as our approximation $\mathbf{W}^k$ contains $\mathbf{V}$, then the algorithm will necessarily select a new direction d^{k+1} that does not "cut into" $\mathbf{V}$, in the sense that $\mathbf{V} \subseteq H(\mathbf{v}^k, d^{k+1})$. Thus, the pivot will orbit around $\mathbf{V}$ in a clockwise manner, so that the sequence of trimmed approximations $\mathbf{W}^k$ will contain $\mathbf{V}$.

The reason is that, under the inductive hypothesis that $\mathbf{V} \subseteq H\left(\mathbf{v}^l, d^l\right)$ for all $l \leq k$, then $\mathbf{V} \subseteq \mathbf{W}^k$ as well, so that $\mathbf{v}^k$, d^k, and $\mathbf{W}^k$ satisfy the assumptions of Lemmas 4 and 5. Thus, as long as $\mathbf{V} \neq \left\{\mathbf{v}^k\right\}$, we know there will exist an action pair a^*, specifically the actions that generate the largest shallowest tangent direction across all states, that generates

a non-binding test direction that is non-zero and shallower than the shallowest test direction. Moreover, if this test direction is not incentive compatible, then we know that the same a^* will generate a binding test direction that is also non-zero and shallower than the shallowest tangent. Thus, a^* generates a regular test direction that is incentive compatible and shallower than the shallowest tangent, and even if this direction is non-monotonic, then we know we can find a continuation direction that is shallower than the shallowest tangent as well.

Since a^* must be considered by our search procedure, we know that D^* is non-empty and contains at least one direction in which payoffs can move without intersection $\mathbf{V}$. As a result, the shallowest test direction d^{k+1} is well-defined and will satisfy the inductive step of containment.

Lemma 6 (Existence)**.** *Suppose that* $\mathbf{V} \subseteq \mathbf{W}^k$ *and* $\mathbf{V} \neq \left\{\mathbf{v}^k\right\}$*. Then there exists a test direction that satisfies* ***P1****–****P4****.*

Lemma 7 (Containment)**.** *Suppose that* $\mathbf{V} \subseteq \mathbf{W}^k$ *and that there exists a test direction that satisfies* ***P1****–****P4****. Then* $\mathbf{V} \subseteq H\left(\mathbf{v}^{k+1}, d^{k+1}\right)$*, so that* $\mathbf{V} \subseteq \mathbf{W}^{k+1}$*.*

We note that if no direction exists that satisfies **P1**–**P4**, then it simply means that either (i) $\mathbf{V} = \left\{\mathbf{v}^k\right\}$, which can be easily verified by checking that $\mathbf{v}^k$ is self-generating, or (ii) if $\mathbf{v}^k(s)$ cannot be generated for some s, then $\mathbf{V}(s) = \emptyset$ and there are no pure strategy subgame perfect Nash equilibria for that state. Recall, however, that we are maintaining the assumption that $\mathbf{V}(s) \neq \emptyset$, to simplify the statements of our subsequent results.

5.4.2 No stalling

Having established that $\mathbf{W}^k$ will not converge to anything smaller than $\mathbf{V}$, we wish to argue that it also cannot converge to anything larger. The key second step in our argument is showing that the algorithm cannot stall, in the sense that starting from any iteration, the pivot will complete a full revolution around $\mathbf{V}$ in finitely many steps.

Let us be more precise about our definition of revolutions. We will refer to a subsequence of iterations of the form $\{l|(r,-1) \leq l \leq (r+1,0)\}$ as a *complete revolution*. Our anti-stalling result is that starting from any k, there exists a $k' > k$ such that the sequence $\{k, \ldots, k'\}$ contains a complete revolution. The logic behind this result is as follows. The algorithm must find a new test direction that satisfies $\mathbf{P}$ at every iteration. If the pivot stopped completing new revolutions around the equilibrium payoff correspondence, then these directions must get stuck at some limit direction, which we denote by d^∞. Thus, for l sufficiently large, $\mathbf{v}^l$ will be strictly increasing in the direction of d^∞.

New test directions can only be generated by three methods: non-binding, binding, and continuation. Moreover, new pivots are always generated as the solution to the system (3) for

some configuration of actions and continuation regimes. Since there are only finitely many states and actions, if the binding payoffs can only take on one of finitely many values, then there are only finitely many ways to configure (3) to generate different pivots. This would be at odds with our hypothesis that there are infinitely many pivots being generated that all increase in the direction d^∞. Thus, there must be infinitely many new binding payoffs being introduced into the system.

Now, recall that the $C(a)$ sets are constant within a revolution. Hence, if the pivot gets stuck and is no longer completing revolutions, the set of continuation values that can be used to generate binding test directions constant as well, and the infinitely many new binding payoffs must be coming from (i) hitting an IC or monotonicity constraint during the pivot update procedure, at which point the regime for that state is changed from non-binding to binding, or (ii) from a continuation direction, in which the pivot travels as far as it can go in the given direction while maintaining monotonicity.

However, (i) or (ii) cannot occur more than once with a given action if d^l is sufficiently close to d^∞. Suppose for the sake of exposition that d^l is exactly d^∞ and is not changing from iteration to iteration. If, say, the best direction at iteration l is a continuation direction generated by action a, then the pivot will travel as far as possible in the direction d^∞ while staying within $\mathbf{W}^k(s) \cap Gen(a)$. This set is a compact and convex polytope that is monotonically decreasing. Thus, if $\mathbf{v}^l(s)$ is already maximized in the direction d^∞, then at subsequent iterations, it will be impossible to move further in this direction using action a. Even if d^l is only very close to d^∞, this will still be true, because eventually d^l will be close enough to d^∞ that moving in any direction between d^l and d^∞ would violate a constraint.

Thus, (i) and (ii) can only happen finitely many times, so that the existence of new directions will eventually require the pivot to complete new revolutions. We therefore have the following:

Lemma 8 (No stalling)**.** *If the algorithm generates infinitely many directions that satisfy* $\boldsymbol{P}$*, then the pivot completes infinitely many revolutions, i.e.,*

$$\lim_{k\to\infty} r(k) = \infty.$$

5.4.3 Convergence

We are almost done. Our algorithm generates a sequence of monotonically decreasing correspondences $\mathbf{W}^k$. It is therefore a consequence of Tarski's fixed point theorem that the $\mathbf{W}^k$ converge to a well-defined limit. The third and last piece of our characterization involves arguing that this limit must be a fixed point of our algorithm. As a consequence, the

limit correspondence must be self-generating, and therefore cannot be strictly larger than $\mathbf{V}$. Since the $\mathbf{W}^k$ are monotonically decreasing, they are converging to a limit

$$\begin{aligned}\mathbf{W}^\infty &= \cap_{k=0}^\infty \mathbf{W}^k. \\ &= \mathbf{W}^0 \cap \left(\cap_{k=0}^\infty H\left(\mathbf{v}^k, d^k\right)\right).\end{aligned}$$

(In the event that at some iteration k the algorithm fails to find a test direction satisfying $\mathbf{P}$, we simply define $\mathbf{W}^l = \left\{\mathbf{v}^k\right\}$ for $l \geq k$, so that $\mathbf{W}^\infty = \left\{\mathbf{v}^k\right\}$.)

It turns out that this limit is self-generating in the sense of APS, and therefore can be no larger than $\mathbf{V}$. Loosely speaking, any payoff v that is in $\mathbf{W}^\infty(s)$ must be in $\mathbf{W}^k(s)$ for all k. As we show in the Appendix, it must be possible to write v as a convex combination of other payoffs that could be generated at the rth revolution from $\mathbf{W}^{r-1:0}$. When new pivots are generated from regular test directions, these payoffs are simply the pivots that are generated on the rth revolution from $\mathbf{W}^{r-1:0}$. The argument is only slightly more subtle when a continuation test direction was used, in which case there are payoffs that can be generated that are "ahead" of the pivot in the direction d^k, which turns out to be sufficient for our purposes. By taking convergent subsequences of the continuation values that generate payoffs whose average is v, we can identify continuation values in the limit set $\mathbf{W}^\infty$ that generate payoffs whose convex combination is v as well, so that $\mathbf{W}^\infty$ self-generates.

Lemma 9 (Self generation)**.** $\mathbf{W}^\infty$ *is self-generating, and therefore is contained in* $\mathbf{V}$*.*

Combining Lemmas 7, 8, and 9, we have our main result:

Theorem 1 (Convergence)**.** *The sequence of approximations* $\{\mathbf{W}^k\}_{k=0}^\infty$ *converges to the equilibrium payoff correspondence, i.e.,* $\cap_{k=0}^\infty \mathbf{W}^k = \mathbf{V}$*.*

6 Application

6.1 A risk sharing example

We will now illustrate our algorithm and methodology by solving a game of informal insurance in the vein of Kocherlakota (1996).[17] Each period, player $i \in \{1, 2\}$ has an endowment of consumption good $e_i \in [0, 1]$, which evolves stochastically over time. The total amount of the good in the economy is constant at $e_1 + e_2 = 1$, so that we can simply write $e = e_1$ and $e_2 = 1 - e$ (cf. Ljungqvist and Sargent, 2004, ch. 20). Thus, e describes the state of the world, which was previously denoted by s.

[17] See also Dixit, Grossman, and Gul (2000), Ligon, Thomas, and Worrall (2000, 2002), and Ljungqvist and Sargent (2004).

The good cannot be stored and must be consumed each period. If player i consumes c_i in a given period, then the flow utility is

$$u(c_i) = \sqrt{c_i}.$$

Note that flow utility is concave, so that players prefer to smooth consumption over time. For example, if players could perfectly smooth their endowment, then the resulting payoffs would be $\sqrt{0.5} \approx 0.705$. On the other hand, if the endowment were independently and uniformly distributed and if a player consumed just their endowment each period, then average utility across states would be only $\int_{x=0}^{1} \sqrt{x} dx \approx 0.667$.

Thus, the players would like to insure themselves against the riskiness of the endowment. While there are no formal insurance contracts, players can make unilateral transfers to subsidize one another's consumption. Let c denote player 1's consumption and let t denote the net transfer from player 1 to player 2. The realized consumption profile is

$$(c, 1-c) = (e-t, 1-e+t).$$

In our subsequent equilibrium analysis, we will implicitly restrict attention to action profiles in which at most one player makes a positive transfer. This is without loss of generality, since any net transfer can be replicated with such an action profile, while at the same time relaxing incentive constraints for both players.

As we have said, the endowment evolves stochastically. We suppose that tomorrow's state e' is distributed according to the density

$$f(e'|c) = \frac{\rho \exp(-\rho |e'-c|)}{2 - \exp(-\rho c) - \exp(-\rho(1-c))}.$$

This density is symmetric around a mode of c with exponential tails that have a shape parameter ρ. As ρ goes to infinity, the distribution converges to a Dirac measure on c, and as ρ goes to zero, the distribution converges to the standard uniform. The density is plotted for various parameter values in Figure 11.

An economic interpretation of these transitions is that each player's productivity is tied to their health and diet, so that players who are well-nourished will, on average, be more productive in the subsequent period. At the same time, there is a fixed amount of resources in the economy, so that one player is well-nourished only if the other player is somewhat malnourished. As an aside, we regard the perfect negative correlation between endowments as somewhat artificial, but it facilitates an apples-to-apples comparisons between economies

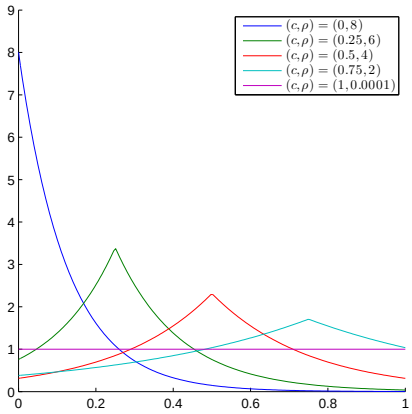


Figure 11: The density $f(e'|c)$ for various values of c and ρ.

with different levels of persistence, since all of these economies have the same feasible payoffs. Thus, the persistence of the endowment affects equilibrium welfare only through incentives.

For the purposes of the computation, both endowment and consumption were taken to be on finite grids

$$e \in E = \left\{0, \frac{1}{K_e - 1}, \ldots, \frac{K_e - 2}{K_e - 1}, 1\right\};$$
$$c \in C = \left\{0, \frac{1}{K_c - 1}, \ldots, \frac{K_c - 2}{K_c - 1}, 1\right\},$$

with K_e and K_c are respectively the number of grid points for endowment and consumption. The consumption grid was chosen to include the endowment grid, so that $K_c = 1 + L(K_e - 1)$ for a positive integer $L \geq 1$. We adapted the continuous probabilities by assigning to each level of the endowment e' the mass in the bin $[e' - 1/(2K_e), e' + 1/(2K_e)]$.

We used our algorithm to compute the equilibria of this game for various discount factors, grid sizes, and levels of persistence ρ. The computations were performed using software that implements our pencil-sharpening algorithm. Specifically, we wrote a library in C++ that we call SGSolve, which contains data structures and routines for representing games and generating the sequence of pivots. The code is general and can be used to solve any game. We note that the algorithm as implemented differs slightly from the algorithm as specified in Section 5, in that the program only generates the regular test directions, which may cause the pivot to move non-monotonically. The program tests a sufficient condition that containment will be satisfied, and it emits a warning if the condition fails.[18] We have also

[18]The key step in our containment argument that relies on monotonicity is showing the existence of a binding test direction that is shallower than the shallowest tangent, in the event that the non-binding test

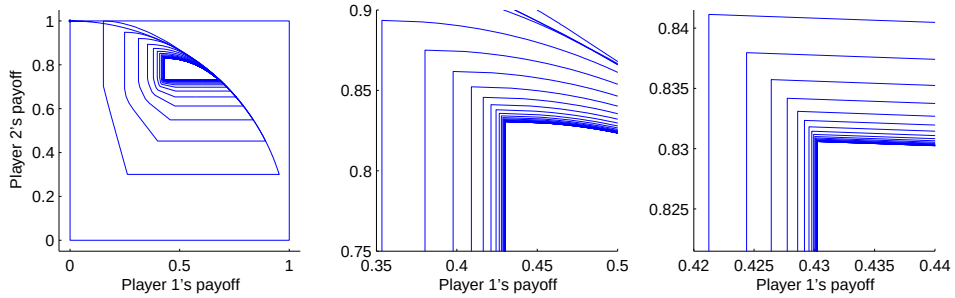


Figure 12: The trajectory of the pivot in state $e = 0$, with $\delta = 0.7$, $K_e = 5$, $K_c = 201$, and $\rho = 0$. The second and third panels show the trajectory around the northwestern corner of $\mathbf{V}(0)$ at greater magnification.

written a graphical interface called SGViewer for interacting with the library. All of these programs and the source code are available through the authors' website,[19] under the terms of the GPLv3 license.[20]

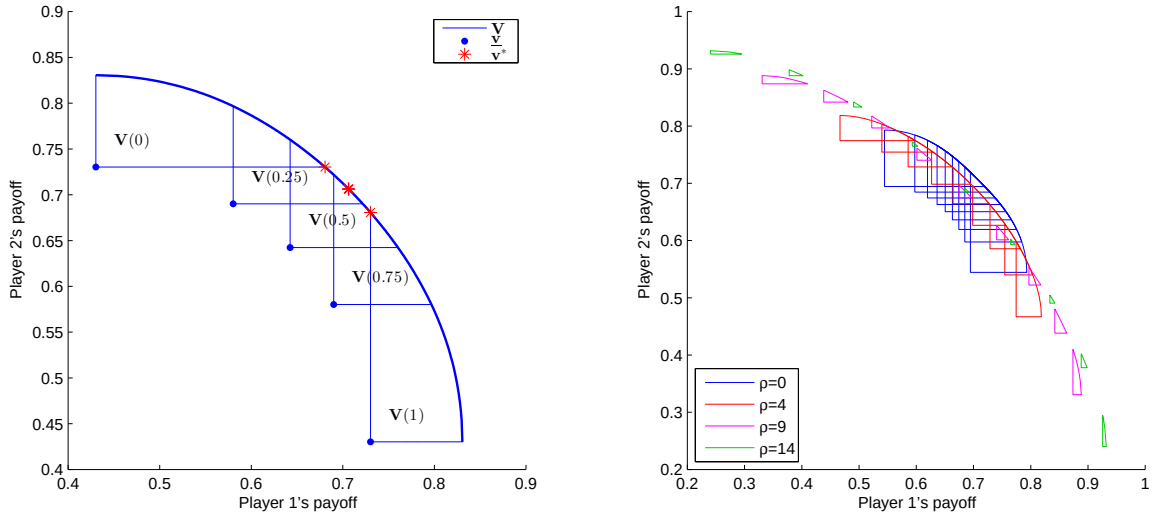
We now describe the computations. Figures 12 and 13(a) present output of the algorithm for the parameter values $\delta = 0.85$, $K_e = 5$, $K_c = 101$, and $\rho = 0$ (so that the endowment is i.i.d. uniform). The algorithm terminated when the distance between successive approximations was less than 10^{-8}, which was attained after 84 revolutions and 52,766 cuts, at a run time of one minute and nine seconds. The final sets have 637 maximal payoff tuples. Figure 12 shows the path taken by the pivot $\mathbf{v}^k(0)$ during the first 20,000 cuts. Figure 13(a) shows the trajectory on the final revolution. The equilibrium payoff sets are outlined in blue and overlap along a northwest-southeast axis. As player 1's endowment e increases from 0 to 1, player 1's payoffs generally increase and payoffs for player 2 generally decrease.

This computation demonstrates key properties of the equilibrium payoff correspondence that are known from prior work, which we will briefly review. Note that the following properties hold for *all* parameter values, not just the ones used for the computation in Figures 12 and 13(a) (e.g., for $\rho > 0$). First, all of the equilibrium payoff sets have right angled southwest frontiers. The corner payoffs, which coincide with the threat tuple $\underline{\mathbf{v}}$, are generated by the "autarkic" equilibrium in which neither player ever makes positive transfers.

direction is not incentive compatible. If the pivot is not feasible, then the shallowest binding test direction may in fact cut into the equilibrium payoff correspondence. However, a sufficient condition for a given binding test direction associated with an action a to not cut into the set is that it is shallower than the slope of the frontier of $Gen(a)$ at the binding payoff that generates the test direction. SGSolve verifies that this is the case whenever a binding test direction is selected as the best direction, and emits a warning if it is not shallower.

[19] www.benjaminbrooks.net/software.shtml

[20] http://www.gnu.org/licenses/gpl-3.0.en.html



(a) (b)

Figure 13: (a) The equilibrium payoff correspondence for $\delta = 0.7$, $K_e = 5$, $K_c = 201$, and $\rho = 0$. (b) Equilibrium payoff correspondences for $\delta = 0.85$, $K_e = 9$, $K_c = 201$, and various values of ρ.

Indeed, this must be the minimum payoff, since a player can always deviate by keeping their endowment for themself, which generates at least the autarky payoff.

Second, the Pareto frontiers of the equilibrium payoff sets all lie along the common frontier indicated by an extra thick blue line. Notice that the evolution of the endowment only depends on current consumption, and not on the current value of the endowment. As a result, the feasible expected continuation value sets, $\overline{V}(c)$, are independent of e. This implies that

$$X(c) = (1-\delta)\, u(c) + \delta \overline{V}(c)$$

is independent of e as well. Moreover, the best deviation is always a transfer of zero, which results in a payoff of exactly $\underline{\mathbf{v}}(e)$. Thus, the set of payoffs that can be generated when the endowment is e is simply

$$X \cap \{v | v \geq \underline{\mathbf{v}}(e)\}$$

where

$$X = \cup_{c \in C} X(c),$$

and the northeastern frontier is simply the Pareto frontier of X.

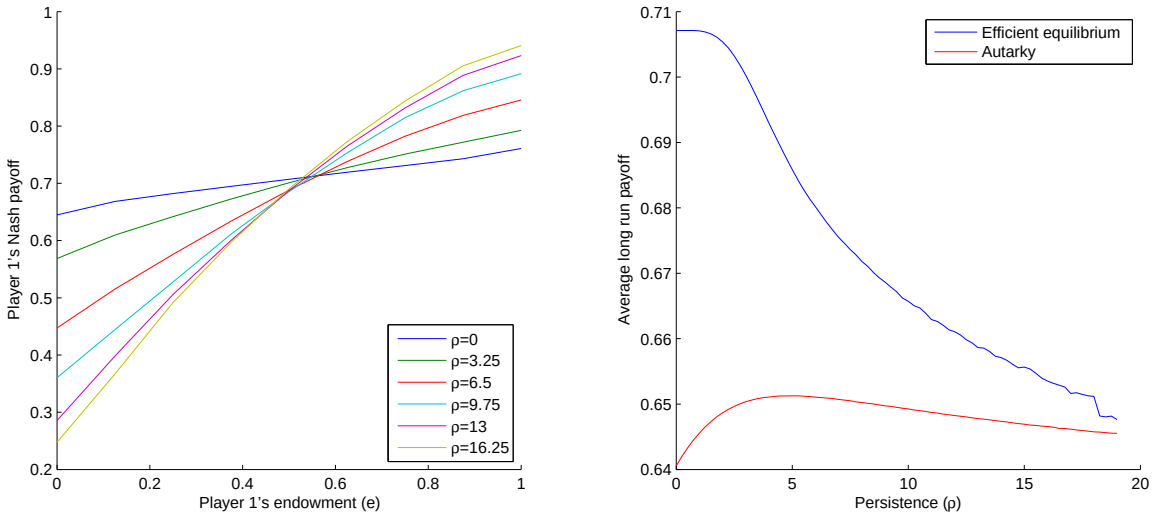
Third, Figure 13(a) allows us to see, in a vivid way, the recursive structure of Section 4. Consider the payoff tuple $\mathbf{v}^*$, depicted with red stars, that maximizes the sum of players' payoffs. Since the Pareto frontiers of $\mathbf{V}(0.25)$, $\mathbf{V}(0.5)$, and $\mathbf{V}(0.75)$ overlap at the 45 degree line, these utilitarian efficient payoffs coincide for these states, i.e., $\mathbf{v}^*(0.25) = \mathbf{v}^*(0.5) = \mathbf{v}^*(0.75)$. Moreover, it must be the same consumption ($c = 0.5$) that generates each of these payoffs. Indeed, since constraints are slack at these levels of the endowment and at this payoff, we know that perfect insurance will obtain until the endowment reaches 0 or 1.

We solved the model for levels of the persistence $\rho \in \{0, 4, 9, 14\}$, with $\delta = 0.85$, $K_e = 9$, and $K_c = 201$. Figure 13(b) displays the equilibrium payoff correspondence for different parameter values. Intuitively, the higher is ρ, the more persistent is the endowment around consumption. This tightens incentive constraints, because when a player deviates by grabbing more consumption today, they also raise their expected endowment in the future. Thus, deviations induce a transition to autarky in a relatively favorable state, thereby weakening the punishment. When ρ equals zero, so that the endowment distribution is always uniformly, it is possible to implement perfect insurance. As ρ increases, the equilibrium payoff sets spread out along the northwest-southeast axis, and even for $\rho = 4$ it is no longer possible to support perfect insurance.

Figure 14 provides two other and complementary visualizations of how payoffs change with the level of persistence. Suppose that the endowment starts at some particular level, and players play the Nash bargaining game to decide which equilibrium should be implemented, where the threat point is the autarky equilibrium. Figure 14(a) shows how the resulting Nash bargaining payoffs depend on the degree of persistence and on the initial state. The results are not terribly surprising: the player with the higher endowment has an advantage in bargaining, and this advantage generally increases as the endowment becomes more persistent.

Now consider a large economy of agents that are engaged in constrained-efficient bilateral risk sharing arrangements. If endowment shocks are independent across pairs, then there is a steady state distribution of consumption in the economy. The blue curve in Figure 14(b) presents the average long run payoffs as a function of ρ. When ρ is close to zero, it is possible to support efficient risk sharing, in which players obtain the efficient surplus of $\sqrt{0.5} \approx 0.705$. As ρ increases, this average payoff declines until risk sharing breaks down altogether for $\rho > 19$.[21] We note that this breakdown occurs somewhat abruptly at high ρ due to the finite consumption grid and discontinuous changes in efficient payoffs when particular discrete transfers can no longer be incentivized.

[21]The current implementation of our algorithm simply stops when there are no directions satisfying $\mathbf{P}$. This may happen because (a) the equilibrium payoff correspondence $\mathbf{V}$ has a single element or (b) there are no pure strategy subgame perfect Nash equilibria, so that $\mathbf{V}$ is empty. In this case, we know that the autarkic equilibrium always exists. Thus, for $\rho > 19$, the efficient and autarky curves would coincide.



(a) (b)

Figure 14: (a) Nash bargaining payoffs and (b) long run payoffs for $\delta = 0.85$, $K_e = 9$, $K_c = 201$, and various values of ρ.

6.2 Computational efficiency

We conclude this section with a comparison to other computational procedures. Because the APS algorithm generates all possible payoffs at each step, the number of extreme points along the APS sequence could in principle grow without bound. In contrast, whenever the monotonicity constraint is not binding, pencil-sharpening has bounded computational complexity per revolution (proportional to the bound on the number of basic tuples).

Judd, Yeltekin, and Conklin (2003), hereafter JYC, proposed an alternative procedure that also has bounded complexity per iteration, at the cost of only approximating the equilibrium correspondence. Fix a finite set of directions

$$D = \{d_1, \ldots, d_{K_d}\}.$$

For a set $W \subset \mathbb{R}^2$, we can define its *outer approximation* to be the smallest convex set with edges parallel to the d_j that contains W. Explicitly, the outer approximation of W is

$$O(W) = \left\{ w | \widehat{d}_j \cdot w \leq b_j \ \forall j = 1, \ldots, K_d \right\},$$

where

$$b_j = \max \left\{ \widehat{d}_j \cdot w | w \in W \right\}.$$

			Run times (h:m:s)							
			$K_d = 100$		$K_d = 200$		$K_d = 400$		$K_d = 600$	
K_e	K_c	#BP	JYC	ABS	JYC	ABS	JYC	ABS	JYC	ABS
3	31	126	1:21.8	1.2						
3	51	203	2:21.2	3.1	7:46.2	3.1				
3	101	395	5:21.9	11.6	16:42.6	11.7	56:50.3	11.5		
5	101	637	13:43	50.7	43:53.2	50.1	2:29:13.1	51.2	5:17:25.5	50.3

Table 2: Run times for various specifications of the risk sharing model and algorithms, in hours:minutes:seconds. #BP denotes the number of basic pairs on the final revolution of pencil sharpening. The convergence criterion for JYC is that distances between sets are less than 10^{-6}, and the convergence criterion for pencil sharpening (ABS) is that the approximation is contained within the final JYC set.

The definition extends to correspondences in the obvious manner: $O(\mathbf{W})(s) = O(\mathbf{W}(s))$.

JYC propose iterating the operator $\widehat{B}(\mathbf{W}) = O(B(O(\mathbf{W})))$, i.e., the outer approximation of the APS set of the outer approximation. Since B and O are both monotonic, $\widehat{B}$ will be monotonic as well. Thus, if $\mathbf{W}^0$ contains $\mathbf{V}$, then so does $\widehat{B}^k(\mathbf{W}^0)$. By taking a rich set of gradients, each $O(\mathbf{W})$ converges to $\mathbf{W}$, so hopefully the limit of this iteration will closely approximate $\mathbf{V}$. Finally, computing $\widehat{B}(\mathbf{W})$ is equivalent to solving a linear programs for each $s \in S$, $a \in A(s)$, and $d \in D$.

For purposes of comparison, we implemented our own version of the JYC algorithm within our C++ libary. For the linear programming, we used the commercial optimization software Gurobi. We also exploited some common structure in the linear programs across j to streamline the computation.

Table 2 reports run times for pencil sharpening and JYC on the risk-sharing example with $\delta = 0.85$, $\rho = 0$, and various grid sizes. In these trials, we first ran the JYC algorithm until the Hausdorff distance between successive approximations was less than 10^{-6}. We then ran pencil sharpening until its approximation was contained in the final JYC set. Thus, the numbers in the "ABS" column represent the time for pencil sharpening to provide a better approximation than JYC. Also, from the output of our algorithm, we know the number of basic pairs that are visited by the pencil-sharpening algorithm on the last revolution, which will generically be the same as the number of extreme points of $\mathbf{V}$.[22] To obtain a comparable

[22]There are two caveats to make here. First, it is a generic possibility that the pencil-sharpening algorithm may take two steps in the same direction; this could occur when the next extreme payoff is generated using a non-binding action a, but the current pivot $\mathbf{v}$ is not incentive compatible for a. In this case, the algorithm may first introduce a into the basic pair with a binding payoff, and then move in the same direction by

level of accuracy, we configured the JYC algorithm with a similar number of directions as the final number of basic pairs, so that both algorithms could in principle generate the same level of detail.

The result are striking. For example, there are 395 basic pairs on the last revolution when $K_e = 3$ and $K_c = 101$, and JYC with 400 gradients takes approximately 56 minutes and 43 seconds to converge. For that same example, pencil sharpening overtakes JYC in 11.5 seconds, which is about 1/300th of the time. Naturally, these numbers should be taken with a grain of salt: there are many ways to implement any given algorithm, and we do not doubt that the implementations of both our own and JYC's algorithm can be refined to reduce run time. Nonetheless, the results strongly suggest that pencil sharpening is significantly faster than existing methods while providing an even greater level of accuracy.

7 Conclusion

It has been the purpose of this paper to study the subgame perfect equilibria of stochastic games. Our work has three distinct components:

(i) We uncover key structural properties of extremal equilibria, namely that extremal equilibrium payoffs can be generated by basic pairs.

(ii) We develop an algorithm that exploits this structure by computing a sequence of basic pairs and corresponding pivot payoffs, and the trajectory of the pivot asymptotically converges to the equilibrium payoff set.

(iii) We implement our algorithm in an accessible format for use by the research community.

AS previously undertook this research program for two player repeated games with perfect monitoring, and we extended the program to the considerably more complex class of stochastic games. The insights that we have used are obviously particular to the two player and perfect monitoring setting. Of course, these are fundamental classes of games both for theory and application and eminently worthy of study. It is our hope that similar efforts will bear fruit for other classes of games, for example, those involving imperfect monitoring or more than two players.

introducing a with a non-binding regime. We regard this as a somewhat exceptional case. Second, on the risk sharing example, the algorithm generates a large number of non-extreme pivots that lie on the "flats" at the southern and western edges of the equilibrium payoff set. This is due to the highly non-generic payoff structure, which causes payoffs generated with binding continuation values to lie along the same line. For a more generic game, each action would generate binding payoffs along different lines. In contrast, all of the basic pairs that are generated along the efficient frontier correspond to distinct extreme points.

References

Abreu, D., D. Pearce, and E. Stacchetti (1986): "Optimal cartel equilibria with imperfect monitoring," *Journal of Economic Theory*, 39, 251–269.

——— (1990): "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, 58, 1041–1063.

Abreu, D. and Y. Sannikov (2014): "An algorithm for two-player repeated games with perfect monitoring," *Theoretical Economics*, 9, 313–338.

Atkeson, A. (1991): "International lending with moral hazard and risk of repudiation," *Econometrica: Journal of the Econometric Society*, 1069–1089.

Blackwell, D. (1965): "Discounted dynamic programming," *The Annals of Mathematical Statistics*, 226–235.

Dixit, A., G. M. Grossman, and F. Gul (2000): "The dynamics of political compromise," *Journal of political economy*, 108, 531–568.

Ericson, R. and A. Pakes (1995): "Markov-perfect industry dynamics: A framework for empirical work," *The Review of Economic Studies*, 62, 53–82.

Hörner, J., T. Sugaya, S. Takahashi, and N. Vieille (2011): "Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a folk theorem," *Econometrica*, 79, 1277–1318.

Judd, K. L., S. Yeltekin, and J. Conklin (2003): "Computing supergame equilibria," *Econometrica*, 71, 1239–1254.

Kocherlakota, N. R. (1996): "Implications of efficient risk sharing without commitment," *The Review of Economic Studies*, 63, 595–609.

Ligon, E., J. P. Thomas, and T. Worrall (2000): "Mutual insurance, individual savings, and limited commitment," *Review of Economic Dynamics*, 3, 216–246.

——— (2002): "Informal insurance arrangements with limited commitment: Theory and evidence from village economies," *The Review of Economic Studies*, 69, 209–244.

Ljungqvist, L. and T. J. Sargent (2004): *Recursive macroeconomic theory*, MIT press.

Mailath, G. J. and L. Samuelson (2006): "Repeated games and reputations: long-run relationships," *OUP Catalogue*.

Phelan, C. and E. Stacchetti (2001): "Sequential equilibria in a Ramsey tax model," *Econometrica*, 69, 1491–1518.

Yeltekin, S., Y. Cai, and K. L. Judd (2015): "Computing equilibria of dynamic games," Tech. rep., Carnegie Mellon University.

A Omitted proofs

A.1 No stalling

The argument for this result will not rely on the particulars of our algorithm, but only on the fact that the pencil-sharpening algorithm generates a trajectory that (1) is monotonically moving in the clockwise direction and (2) contains the non-empty convex sets $\mathbf{V}(s)$. These properties are formalized as:

1. $\mathbf{v}^l(s) = \mathbf{v}^{l-1}(s) + \mathbf{x}^l(s)d^l$ where $d^l \cdot \widehat{d}^{l-1} \leq 0$ and $\mathbf{x}^l(s) \in \mathbb{R}_+$.
2. $v \cdot \widehat{d}^l \leq \mathbf{v}^l(s) \cdot \widehat{d}^l$ for all $v \in \mathbf{V}(s)$.

We will refer to such a sequence $\left\{\mathbf{v}^l, d^l\right\}$ as an *orbital trajectory*. The containment argument proves that the algorithm generates an orbital trajectory.

We will say that the orbital trajectory $\left\{\mathbf{v}^k, d^k\right\}$ has the *rotation property* if for all directions d and for all k, there exists an $l \geq k$ such that $d^l \cdot \widehat{d} \leq 0$. In words, this property says that for every direction, the sequence will point weakly below that direction infinitely many times.

We will argue that the trajectory generated by the twist algorithm satisfies the rotation property, which in turn implies that the algorithm completes infinitely many revolutions. The proof is lengthy, but the intuition is quite straightforwrd. If the pivot ever got stuck in a way that prevented it from completing orbits of $\mathbf{V}$, then eventually the direction would move around and cut some payoffs in $\mathbf{V}$ out of the trajectory, which contradicts containment.

Lemma 10. *If the orbital trajectory $\left\{\mathbf{v}^k, d^k\right\}$ satisfies the rotation property, $\lim_{k\to\infty} r(k) = \infty$.*

Proof of Lemma 10. We will show that from any iteration k, there exists an iteration $k' > k$ that starts a revolution, which implies the result.

First, for any k, there must exist a $k' > k$ such that $d^{k'} \cdot \widehat{d}^k < 0$. The rotation property would clearly fail for the direction $-\widehat{d}^k$ if $d^l \propto d^k$ for all $l > k$, and if we only have $d^l \propto d^k$

or $d^l \propto -d^k$, then containment would be violated under the hypothesis that $\mathbf{V}$ has full dimension.

Now let us consider two cases. If $d^k \propto d^N$, then there exists a smallest $k' > k$ such that $d^{k'} \cdot \widehat{d}^k = d^{k'} \cdot \widehat{d}^N < 0$, which in fact must start a revolution.

Otherwise, if $d^k \cdot \widehat{d}^N > 0$, there is a smallest $k_1 \geq k$ such that $d^{k_1} \cdot \widehat{d}^N > 0$. There is then a smallest $k_2 \geq k_1$ such that $d^{k_2} \cdot \widehat{d}^N \leq 0$, and finally a smallest $k_3 \geq k_2$ such that $d^{k_3} \cdot \widehat{d}^N < 0$. We claim that k_3 starts a revolution. If $d^{k_3-1} \cdot \widehat{d}^N > 0$, then this is obvious. Otherwise, we claim that $d^{k_3-1} \propto d^N$. For if $d^{k_3-1} \propto -d^N$, then $d^{k_3} \cdot \widehat{d}^{k_3-1} = -d^{k_3} \cdot \widehat{d}^N > 0$, a contradiction. □

Lemma 11. *If the rotation property fails, then there exists a direction d^∞ such that $d^l/\|d^l\| \to d^\infty$, and moreover $d^l \cdot \widehat{d}^\infty \geq 0$ for l sufficiently large.*

Proof of Lemma 11. Suppose that there exists a k and direction $\underline{d}$ such that $d^l \cdot \widehat{\underline{d}} > 0$ for all $l \geq k$. We can write each direction d^l as

$$d^l/\|d^l\| = x^l \underline{d} + y^l \widehat{\underline{d}}$$

for some coordinates x^l and y^l. Note that the hypothesis $d^l \cdot \widehat{\underline{d}} > 0$ implies that $y^l > 0$.

Claim: x^l is monotonically increasing in l. The best direction d^l must satisfy $d^l \cdot \widehat{d}^{l-1} \leq 0$, which implies that

$$\begin{aligned} d^l \cdot \widehat{d}^{l-1} &= (x^l \underline{d} + y^l \widehat{\underline{d}})(x^{l-1} + y^{l-1} \widehat{\underline{d}}) \\ &= (x^{l-1} y^l - x^l y^{l-1}) \|\underline{d}\|^2 \leq 0 \end{aligned}$$

so that

$$x^{l-1} y^l \leq x^l y^{l-1}.$$

Suppose that $x^l < x^{l-1}$. Then $y^l > y^{l-1}$ (since $(x^l)^2 + (y^l)^2 = 1$), so

$$x^l y^{l-1} < x^l y^l < x^{l-1} y^l$$

since $y^l > 0$, a contradiction. Thus, it must be that $x^l > x^{l-1}$. It must also be that $x^l \leq 1$, so that x^l converges to some x^∞. Finally, $y^l = \sqrt{1 - (x^l)^2}$, so y^l converges to $y^\infty = \sqrt{1 - (x^\infty)^2}$, and the limit direction is

$$d^\infty = x^\infty \underline{d} + y^\infty \widehat{\underline{d}}.$$

In the limit, $d^l \cdot \widehat{d^\infty}$ is proportional to $x^\infty y^l - x^l y^\infty$. Monotonicity implies that $x^\infty \geq x^l$ and x^∞ and x^l have the same sign. Thus, if $x^\infty > 0$, x^l must be positive so that $y^l \geq y^\infty$, so that $x^\infty y^l \geq x^l y^\infty$. If $x^l \leq x^\infty \leq 0$, then $y^l \leq y^\infty$, and again we conclude that $x^\infty y^l \geq x^l y^\infty$. □

Having established these general results about orbital trajectories, we can now return to the particulars of our algorithm and prove the anti-stalling lemma.

Proof of Lemma 8. Suppose that the trajectory generated by the algorithm does not complete infinitely many revolutions. Then from Lemma 10, we conclude that the rotation property fails, so that there exists a k and a $\underline{d}$ such that $d^l \cdot \widehat{\underline{d}} \geq 0$ for all $l \geq k$. We then conclude from Lemma 11 there exists a direction d^∞ such that $d^l/\|d^l\| \to d^\infty$. Moreover, there exists a k' such that for all $l \geq k'$, $d^l \cdot \widehat{d^\infty} \geq 0$ and $d^l \cdot d^\infty > 0$. We also note that there must exist a k' such that no new revolutions are completed after iteration k'. In particular, if d^∞ points west of due north, then eventually all of the d^l will point west of due north, so that it will be impossible for d^l to point east again and no new revolutions can be completed. The analysis would be symmetric if d^∞ points east. Thus, we can assume without loss of generality that the sets $\overline{W}(a)$, $IC(a)$, $Gen(a)$, $C(a)$, and the function $h(a)$ are not changing.

Because there are finitely many actions and states, there must be some action which generates infinitely many best test directions, and for l sufficiently large, we must have that $\mathbf{v}^l(s)$ is strictly increasing in the d^∞ direction. Now, there are only finitely many binding payoffs in the pivot $\mathbf{v}^k$ and only finitely many extreme continuation values in $C(a)$. As a result, there are only finitely many configurations of the system that defines $\mathbf{v}^l$ that use (i) binding payoffs that were in the original pivot $\mathbf{v}^k$ or (ii) extreme binding continuation values in $C(a)$. Thus, it is impossible that the algorithm generates infinitely many new pivots that are monotonically increasing in the d^∞ direction using only non-binding and binding payoffs.

There are therefore only two possibilities for generating new directions, by introducing new binding payoffs into the pivot: (i) new binding payoffs introduced when changing a non-binding regime to a binding regime in state s during the pivot updating procedure or (ii) generating a new continuation test direction.

Now let us consider two cases. In the first case, there exists a k such that for all $l \geq k$, $d^l/\|d^l\| = d^\infty$, i.e., d^l converges exactly in finitely many iterations. Now, it is not too hard to see that this is incompatible with a generating infinitely many non-zero movements according to (i) or (ii). For example, when (i) occurs, the pivot must travel as far as possible in the direction d^∞ while maintaining incentive compatibility. If $\mathbf{v}^l(s)$ were to travel any further in the direction d^∞ on a subsequent iteration using the same action, incentive compatibility would be violated, which rules out new pivots of the forms (i), (ii), or (iii) being generated with this action. Similarly, if a new pivot were generated according to (ii), the pivot again

moves as far as possible in the direction d^∞ subject to the continuation value being (a) incentive compatible, (b) feasible in the sense of $w \in \overline{W}(a)$, and (c) contained in $\mathbf{W}^k(s)$ (which contains $\mathbf{W}^l(s)$ for all $l > k$). Thus, any further movement in the direction d^∞ on a subsequent iteration must violate one of these requirements (a-c), and therefore is impossible.

In the second case, d^l approaches d^∞ in the limit but never converges exactly. Let

$$X(a) = Gen(a) \cap \mathbf{W}^k(s)$$

denote the payoffs in $\mathbf{W}^k$ that can be generated using a and feasible and incentive compatible continuation values in state s. The set $X(a)$ is a convex polytope in $\mathbb{R}^2$. Let M denote the set of directions which are the slopes of edges of $X(a)$. In other words, if E is an edge of $X(a)$, then

$$E \subseteq \overline{H}\,(v, m)$$

for some $v \in X(a)$ and $m \in M$, where

$$\overline{H}(v, m) = \{w | w \cdot \widehat{m} = v \cdot \widehat{m}\}$$

is the line through v with slope m. Since there are only finitely many pivots up to iteration k, $X(a)$ has finitely many extreme points, and therefore M is a finite set. Let k' be large enough so that (i) $d^{k'} \cdot \widehat{m} \neq 0$ for all $m \in M$, and (ii) $\text{sgn}\left(d^{k'} \cdot \widehat{m}\right) = \text{sgn}\left(d^l \cdot \widehat{m}\right)$ for all $l \geq k'$. This k' must exist beause d^l converges. For example, if $d^\infty \cdot \widehat{m} > 0$, then obviously there exists a k_m so that for all $l \geq k_m$, $d^l \cdot \widehat{m} > 0$ as well. This will symmetrically be true for $d^\infty \cdot \widehat{m} < 0$. If $d^\infty \cdot \widehat{m} = 0$, then we can establish the existence of k_m so that if $d^\infty \cdot m > 0$ (< 0), then there exists a k_m such that $d^l \cdot \widehat{m} > 0$ (< 0). For in the former case, $d^\infty = xm$ for some $x > 0$, so $d^l \cdot \widehat{m} = d^l \cdot x\widehat{d^\infty}$, which is strictly positive. In the other case, we use the fact that $d^\infty = -xm$ for some $x > 0$.

Now let

$$Y^l = \left\{w | w \cdot \widehat{d^l} \leq \mathbf{v}^l(s) \cdot \widehat{d^l}, w \cdot \widehat{d^\infty} > \mathbf{v}^l(s) \cdot \widehat{d^\infty}\right\}$$

be the set of payoffs which could generate a new direction d^{l+1} such that $d^{l+1} \cdot \widehat{d^l} \leq 0$ and $d^{l+1} \cdot \widehat{d^\infty} > 0$. Note that $Y^l \subseteq Y^{l-1}$. Any payoff $w \in Y^l$ can be written as

$$\begin{aligned} w &= xd^l + yd^\infty; \\ \mathbf{v}^l(s) &= x^l d^l + y^l d^\infty; \\ \mathbf{v}^{l-1}(s) &= x^{l-1} d^l + y^{l-1} d^\infty. \end{aligned}$$

The fact that $w \cdot \widehat{d^\infty} > \mathbf{v}^l(s) \cdot \widehat{d^\infty}$ implies that $x > x^l$, since $d^l \cdot \widehat{d^\infty} > 0$. In turn, $\mathbf{v}^l(s) \in Y^{l-1}$ implies that $\mathbf{v}^l(s) \cdot \widehat{d^\infty} > \mathbf{v}^{l-1}(s) \cdot \widehat{d^\infty}$ and therefore $x^l \geq x^{l-1}$, which proves that $x \geq x^{l-1}$. On the other hand,

$$\begin{aligned} w \cdot \widehat{d^l}^{l-1} &= xd^l \cdot \widehat{d^l}^{l-1} + yd^\infty \cdot \widehat{d^l}^{l-1} \\ &\leq x^l d^l \cdot \widehat{d^l}^{l-1} + y^l d^\infty \cdot \widehat{d^l}^{l-1} \end{aligned}$$

since $w \cdot \widehat{d^l} \leq \mathbf{v}^l(s) \cdot \widehat{d^l}$ implies that $y \leq y^l$, as $d^\infty \cdot \widehat{d^l} < 0$, and $d^l \cdot \widehat{d^l}^{l-1} \leq 0$ as well. The latter implies that $\mathbf{v}^l(s) \cdot \widehat{d^l}^{l-1} \leq \mathbf{v}^{l-1}(s) \cdot \widehat{d^l}^{l-1}$, so that $w \cdot \widehat{d^l}^{l-1} \leq \mathbf{v}^{l-1}(s) \cdot \widehat{d^l}^{l-1}$.

Now, let $k'' \geq k'$ such that a generates the best continuation direction. (The analysis for case (i) is entirely analogous, with the incentive constraints replacing the half-space constraints that define $X(a)$.) This implies that $\mathbf{v}^{k''}(s)$ is on the boundary of $X(a)$, and in particular that $\mathbf{v}^{k''}(s)$ is on an edge E with slope m.

Claim: $Y^{k''} \cap X(a) = \emptyset$. Note that any $\mathbf{v}^l(s)$ for $l \geq k''$ must be contained in $Y^l \cap X(a)$, which is contained in $Y^{k''} \cap X(a)$. Thus, a consequence of this claim is that a cannot generate any more non-zero directions at iterations later than k'' as we had supposed.

Now, let us see why the claim is true. Note that $X(a) \subseteq H\left(\mathbf{v}^{k''}(s), m\right)$ for some $m \in M$, which is the slope of an edge that contains $X(a)$. If $\mathbf{v}^l(s)$ is not an extreme point, this m is unique, and we note that it must be the case that $d^{k''-1} \cdot \widehat{m} > 0$. Otherwise, traveling further in the direction $d^{k''-1}$ would move the pivot into the interior of $X(a)$, contradicting that we had moved as far as possible in the direction $d^{k''-1}$ without violating feasibility or incentive compatibility.

On the other hand, if $\mathbf{v}^{k''}(s)$ is an extreme point, there are two such m. We can distinguish these as m^1 and m^2, where m^2 is the slope of the clockwise edge and m^1 is the slope of the counter-clockwise edge. We claim that for at least one of these m, it must be that $d^{k''-1} \cdot \widehat{m} > 0$. Otherwise, the same claim applies as above. In particular, if $d^{k''-1} = xm^2$ or if $d^{k''-1} = -xm^1$ for some $x > 0$, then it is clearly possible to move along one of the edges. If $d^{k''-1} = -xm^2$ or if $d^{k''} = xm$, then because $m^2 \cdot \widehat{m}^1 < 0$, either $d^{k''-1} \cdot \widehat{m}^1 < 0$

or $d^{k''-1} \cdot \widehat{m}^2 < 0$. Finally, if $d^{k''-1} \cdot \widehat{m}^1 > 0$ and $d^{k''-1} \cdot \widehat{m}^2 > 0$, then by traveling in the direction $d^{k''-1}$, the pivot would move into the interior of $X(a)$.

Thus, we can find an m for which $X(a) \subseteq H\left(\mathbf{v}^{k''}(s), m\right)$ and $d^{k''-1} \cdot \widehat{m} < 0$. This implies that $d^l \cdot \widehat{m} > 0$ for all $l \geq k''$, and in particular, that $d^\infty \cdot \widehat{m} \geq 0$. It will be sufficient to show that for all $w \in Y^l$, $w \cdot \widehat{m} > \mathbf{v}^{k''}(s) \cdot \widehat{m}$. Let us write

$$\begin{aligned} w &= x d^{k''-1} + y d^\infty \\ \mathbf{v}^l(s) &= x^l d^{k''-1} + y^l d^\infty. \end{aligned}$$

Then

$$\begin{aligned} w \cdot \widehat{d^\infty} &= x d^{k''-1} \cdot \widehat{d^\infty} \\ &> x^{k''} d^{k''-1} \cdot \widehat{d^\infty} \\ &= \mathbf{v}^{k''}(s) \cdot \widehat{d^\infty}, \end{aligned}$$

which implies that $x > x^{k''}$, since $d^{k''-1} \cdot \widehat{d^\infty} > 0$. Similarly,

$$\begin{aligned} w \cdot \widehat{d^{k''-1}} &= y d^\infty \cdot \widehat{d^{k''-1}} \\ &\leq y^{k''} d^\infty \cdot \widehat{d^{k''-1}} \\ &= \mathbf{v}^{k''}(s) \cdot \widehat{d^{k''-1}}, \end{aligned}$$

which implies that $y \geq y^{k''}$, since $d^\infty \cdot \widehat{d^{k''-1}} < 0$. Thus, since $d^\infty \cdot \widehat{m} > 0$ and $d^{k''-1} \cdot \widehat{m} \geq 0$, we conclude that

$$\begin{aligned} w \cdot \widehat{m} &= x d^{k''-1} \cdot \widehat{m} + y d^\infty \cdot \widehat{m} \\ &> x^{k''} d^{k''-1} \cdot \widehat{m} + y^{k''} d^\infty \cdot \widehat{m} \\ &= \mathbf{v}^{k''}(s) \cdot \widehat{m}, \end{aligned}$$

so $w \notin H\left(\mathbf{v}^{k''}, m\right)$ and $w \notin X(a)$. □

A.2 Convergence

To prove convergence, we need another "purely geometric" fact about orbital trajectories. Let us say that the subsequence $\left\{\left(\mathbf{v}^l, d^l\right)\right\}_{l=k'}^{k''}$ has the *covering property* if for all $d \in \mathbb{R}^2$, there exist $l \in \{k', \ldots, k''-1\}$ and $\alpha, \beta \geq 0$ such that

$$d = \alpha d^l + \beta d^{l+1}.$$

In other words, d lies between d^l and d^{l+1}. The first result is the following:

Lemma 12 (Covering)**.** *Suppose the subsequence* $\left\{\left(\mathbf{v}^l, d^l\right)\right\}_{l=k'}^{k''}$ *satisfies the covering property, and let* $\left\{\mathbf{x}^l\right\}_{l=k'}^{k''}$ *be tuples of non-negative scalars. Then*

$$\cap_{l=k'}^{k''} H\left(\mathbf{v}^l, d^l\right) \subseteq co\left(\cup_{l=k'}^{k''}\left\{\mathbf{v}^l + \mathbf{x}^l d^l\right\}\right).$$

Proof of Lemma 12. Let

$$X = \cap_{l=k'}^{k''-1} H\left(\mathbf{v}^l(s), d^l\right),$$

and let

$$Y = \mathrm{co}\left(\cup_{l=k'}^{k''}\left\{\mathbf{v}^l(s) + \mathbf{x}^l\left(s\right) d^l\right\}\right)$$

denote the convex hull of the trajectory of the subsequence in state s, which are both convex sets.

Suppose that there exists a $v \in X \setminus Y$. By the separating hyperplane theorem, there is a direction d such that $w \cdot \widehat{d} < v \cdot \widehat{d}$ for all $w \in Y$. In particular, $\left(\mathbf{v}^l(s) + \mathbf{x}^l\left(s\right) d^l\right) \cdot \widehat{d} < v \cdot \widehat{d}$ for all $l = k', \ldots, k''$. Because of the covering property, we can find an $l \in \{k', \ldots, k'' - 1\}$ and $\alpha, \beta \geq 0$ such that $d = \alpha d^l + \beta d^{l+1}$. Note that $v \in X$ implies that $v \in H\left(\mathbf{v}^l(s), d^l\right)$ and also that $v \in H\left(\mathbf{v}^{l+1}(s), d^{l+1}\right) = H\left(\mathbf{v}^l(s), d^{l+1}\right)$. This means that

$$\begin{aligned} v \cdot \widehat{d^l} &\leq \mathbf{v}^l(s) \cdot \widehat{d^l} = \left(\mathbf{v}^l\left(s\right) + \mathbf{x}^l\left(s\right) d^l\right) \cdot \widehat{d^l}; \\ v \cdot \widehat{d^{l+1}} &\leq \mathbf{v}^l(s) \cdot \widehat{d^{l+1}} \leq \left(\mathbf{v}^l(s) + \mathbf{x}^l\left(s\right) d^l\right) \cdot \widehat{d^{l+1}}, \end{aligned}$$

due to the fact that $d^l \cdot \widehat{d^{l+1}} \geq 0$, so that

$$\begin{aligned} v \cdot \widehat{d} &= \alpha v \cdot \widehat{d^l} + \beta v \cdot \widehat{d^{l+1}} \\ &\leq \alpha\left(\mathbf{v}^l(s) + \mathbf{x}^l d^l\right) \cdot \widehat{d^l} + \beta\left(\mathbf{v}^l(s) + \mathbf{x}^l d^l\right) \cdot \widehat{d^{l+1}} \\ &= \left(\mathbf{v}^l(s) + \mathbf{x}^l d^l\right) \cdot \widehat{d}, \end{aligned}$$

so that d cannot in fact separate v from Y. □

Naturally, we will need the fact that complete revolutions of the pivot satisfy the covering property.

Lemma 13 (Complete revolutions)**.** *A complete revolution satisfies the covering property.*

Proof of Lemma 13. Let $d \in \mathbb{R}^2$ be an arbitrary direction, and let us suppose that $d \cdot \widehat{d^N} \leq 0$. The case where $d \cdot \widehat{d^N} \geq 0$ is symmetric and is omitted.

We will show that $d \in X^l$ for some l, where

$$X^l = \left\{\alpha d^l + \beta d^{l+1} | \alpha \geq 0, \beta \geq 0\right\}$$

for $l = k, \ldots, k' - 1$. Note that the X^l are additive (convex) cones, and some of the X^l may be one-dimensional if $d^{l+1} = x d^l$ for some $x > 0$. Note that a sufficient condition for d to be in X^l, as long as $d^l \not\propto d^{l+1}$, is that $d \cdot \widehat{d^{l+1}} \leq 0$ and $d \cdot \widehat{d^l} \geq 0$. This can be easily verified by decomposing d in (d^l, d^{l+1}) coordinates and verifying that the coefficients are both positive.

Now, observe that X^k contains d^N. Since $d^{k+1} \cdot \widehat{d^N} < 0$, there is a smallest $\widetilde{k} \in \{k+2, \ldots, k'\}$ such that $d^{\widetilde{k}} \cdot \widehat{d^N} \geq 0$, which we know exists because k' starts a revolution so it is true for $\widetilde{k} = k' - 1$.

Claim: $-d^N \in X^{\widetilde{k}-1}$. Why? $d^{\widetilde{k}-1} \cdot \widehat{d^N} < 0$ and $d^{\widetilde{k}} \cdot \widehat{d^N} \geq 0$, so that there exists some positive weights α and β such that $d' = \alpha d^{\widetilde{k}-1} + \beta d^{\widetilde{k}}$ satisfies $d' \cdot \widehat{d^N} = 0$. By scaling up or down, we can ensure that either $d' = d^N$ or $d' = -d^N$. Moreover, we know that $d' \cdot \widehat{d^{\widetilde{k}-1}} \leq 0$, which cannot be true if $d' = d^N$.

Thus, we can define the following sets:

$$\begin{aligned} Y^k &= \left\{\alpha d^N + \beta d^{k+1} | \alpha \geq 0, \beta \geq 0\right\}; \\ Y^{\widetilde{k}-1} &= \left\{\alpha d^{\widetilde{k}-1} + \beta(-d^N) | \alpha \geq 0, \beta \geq 0\right\}; \\ Y^l &= X^l \text{ if } k < l < \widetilde{k} - 1. \end{aligned}$$

Suppose that $d \notin X^l$ for any l. Then since $Y^l \subseteq X^l$ for $l = k, \ldots, \widetilde{k} - 1$, we conclude that $d \notin Y^l$ either. We shall see that this leads to a contradiction.

In particular, since $d \cdot \widehat{d^N} \leq 0$, $d \notin Y^k$ implies that $d \cdot \widehat{d^{k+1}} < 0$. Continuing inductively for $l = k+1, \ldots, \widetilde{k} - 2$, if $d \cdot \widehat{d^l} < 0$ and $d \notin Y^l = X^l$, then we conclude that $d \cdot \widehat{d^{l+1}} < 0$. Once we reach $l = \widetilde{k} - 1$, we know that $d \cdot \widehat{d^{\widetilde{k}-1}} < 0$, and since $d \notin Y^{\widetilde{k}-1}$, we conclude that $-d \cdot \widehat{d^N} < 0$, or equivalently that $d \cdot \widehat{d^N} > 0$, a contradiction. □

Proof of Lemma 9. The result is trivial if $\mathbf{W}^\infty(s)$ is a singleton, in which case containment and the assumption that $\mathbf{V}(s)$ is non-empty imply that $\mathbf{V}(s) = \mathbf{W}^\infty(s)$. Similarly, we can dispense with the case in which the algorithm fails to find a direction satisfying $\mathbf{P}$, in which case again $\mathbf{V} = \left\{\mathbf{v}^k\right\}$. Thus, Lemma 8 implies that the algorithm completes infinitely many revolutions.

We will argue that every $v \in \mathbf{W}^\infty(s)$ is a convex combination of payoffs that can be generated by some $a \in \mathbf{A}(s)$ and $\mathbf{w} \in \mathbf{W}^\infty$. Since $\mathbf{W}^k$ is monotonically decreasing, it must

be that $v \in \mathbf{W}^{r:0}(s)$ for all revolutions r. Recall that a pivot payoff $\mathbf{v}^k(s)$ may have been generated in one of three ways: (i) non-binding, (ii) binding, or (iii) continuation. In the first two cases, $\mathbf{v}^k(s)$ is generated by some actions a and continuation payoffs $\mathbf{w} \in \mathbf{W}^{r(k):0}$. In case (iii), however, $\mathbf{v}^k(s)$ is not necessarily generated, but rather it may be the convex hull of $\mathbf{v}^{k-1}(s)$ and a payoff $\mathbf{v}^k(s) + x^k d^k \in Gen\left(\mathbf{a}^k(s)\right)$ with $x^k \geq 0$ that lies along the ray extending from $\mathbf{v}^{k-1}(s)$ in the clockwise tangent direction to $\mathbf{W}^{k-1}$ (which by assumption is d^k). Let us also define $x^k = 0$ for iterations k in which the pivot was generated via (i) or (ii), and define

$$Y^r = \text{co}\left(\cup_{l=r:0}^{r+1:0}\left\{\mathbf{v}^l(s) + x^l d^l\right\}\right).$$

Lemmas 12 and 13 show that $\mathbf{W}^{r:0} \subseteq Y^{r+1}$ for each r.

Since $v \in \mathbf{W}^{r:0}$, we can write v as a convex combination of at most three extreme points of Y^{r+1}:

$$v = \sum_{l=1}^{3} \alpha^{r,l} \mathbf{v}^{r,l}(s),$$

where $\mathbf{v}^{r,l}(s) \in Y^{r+1}$ for $l = 1, 2, 3$. Now, each of these pivots must have been generated from actions $a^{r,l} \in \mathbf{A}(s)$ and continuation values $\mathbf{w}^{r,l} \in \mathbf{W}^{r-1:0}$. Thus,

$$v = \sum_{l=1}^{3} \alpha^{r,l} \left((1-\delta) g\left(a^{r,l}\right) + \delta \sum_{s' \in S} \pi\left(s' | a^{r,l}\right) \mathbf{w}^{r,l}(s') \right)$$

and for each r, l, i, and $a_i' \in \mathbf{A}_i(s)$, we must have

$$\mathbf{v}_i^{r,l}(s) \geq (1-\delta) g_i\left(a_i', a_{-i}^{r,l}\right) + \delta \sum_{s' \in S} \pi\left(s' | a_i', a_{-i}^{r,l}\right) \underline{\mathbf{w}}^r(s'),$$

where $\underline{\mathbf{w}}^r = \underline{\mathbf{w}}\left(\mathbf{W}^{r:0}\right)$. Each of the sequences $\left\{a^{r,l}\right\}$, $\left\{\alpha^{r,l}\right\}$, $\left\{\mathbf{w}^{r,l}\right\}$, and $\left\{\underline{\mathbf{w}}^r\right\}$ lie in compact metric spaces $\mathbf{A}(s)$, $[0,1]$, and $\mathbf{W}^0$ respectively, and so there is a subsequence for which all of these objects converge to limits a^l, α^l, $\mathbf{w}^l$, and $\underline{\mathbf{w}}$ respectively. Moreover, it must be that $\mathbf{w}^l \in \mathbf{W}^\infty$ and $\underline{\mathbf{w}} \in \mathbf{W}^\infty$ for all $l = 1, 2, 3$. We argue this for $\mathbf{w}^l$: For every k, there is an r such that $r : 0 > k$, and from this point on,

$$\mathbf{w}^{r,l} \in \mathbf{W}^{r:0} \subseteq \mathbf{W}^k \subseteq H\left(\mathbf{v}^k, d^k\right).$$

Since $H\left(\mathbf{v}^k, d^k\right)$ is closed, $\mathbf{w}^{r,l}$ must converge to a point in $H\left(\mathbf{v}^k, d^k\right)$. But this is true for every k, thus proving that $\mathbf{w}^l \in H\left(\mathbf{v}^k, d^k\right)$ for all k. Moreover, $\mathbf{w}^{r,l} \in \mathbf{W}^0$ for every r and l, and $\mathbf{W}^0$ is also closed, so that $\mathbf{w}^l \in \mathbf{W}^0$ as well, and we conclude that $\mathbf{w}^l \in \mathbf{W}^\infty$.

It is now obvious that

$$v = \sum_{l=1}^{3} \alpha^l \left((1-\delta)\, g\left(a^l\right) + \delta \sum_{s' \in S} \pi\left(s'|a^l\right) \mathbf{w}^l\left(s'\right) \right)$$

and

$$\begin{aligned} \mathbf{v}_i^l(s) &\geq (1-\delta)\, g_i\left(a_i', a_{-i}^l\right) + \delta \sum_{s' \in S} \pi\left(s'|a_i', a_{-i}^l\right) \underline{\mathbf{w}}\left(s'\right) \\ &\geq (1-\delta)\, g_i\left(a_i', a_{-i}^l\right) + \delta \sum_{s' \in S} \pi\left(s'|a_i', a_{-i}^l\right) \underline{\mathbf{w}}\left(\mathbf{W}^\infty\right)\left(s'\right), \end{aligned}$$

thus proving that v is a convex combination of payoffs that can be generated by $\mathbf{W}^\infty$. $\square$