Approximating Markov Perfect Equilibria for Nonstationary Dynamic Games

Alfredo Garcia
Robert L. Smith
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109

Technical Report 96-16

November 1996

Approximating Markov Perfect Equilibria for Nonstationary Dynamic Games

Alfredo Garcia and Robert L. Smith
Department of Industrial and Operations Engineering
University of Michigan Ann Arbor, MI 48109

1 Introduction

A sequential characterization[1] of infinite horizon subgame perfect Nash equilibria as limits of Finite Horizons subgame perfect Nash equilibria is not possible due to "end of horizon" effects (see, for instance, Fudenberg and Levine[1983] and [1986]). This can be illustrated, with the celebrated game "prisoners dilemma" where "always defect" is the unique equilibrium outcome of any finite horizon repetition of the game. However, when the game is repeated *ad infinitum* then, "always cooperate" is the equilibrium outcome of a certain Perfect Nash equilibrium for a certain range of discount rates. This situation is due to the fact that for a fixed finite horizon, "defection" in the last time period dominates "cooperation", hence "always cooperate" can not be an equilibrium outcome for such fixed horizon.

The first work attempted to overcome this phenomenon was carried out by Fudenberg and Levine[1983].They build on the work by Radner[1980], who pointed out that if players are boundedly rational, or equivalently, if one relaxes the definition of equilibria up to an epsilon, making epsilon smaller as the horizon diverged, then the set of epsilon-equilibria of finite horizons converged to the set of infinite horizon equilibria. Fudenberg and Levine's work focused on the space of strategies, in which, they define a suitable metric topology, that facilitates the result.

Pearce[1990], explains the relationship of Radner's and Fudenberg and Levine's work to the *self-generation* approach used by Abreu, Pearce and Stacchetti [1986] and [1990] for repeated games. They exploit the structure of repeated games with imperfect monitoring to obtain a

[1] see Klein et al. [1984] for the Kuratowsky definition of set convergence.

dynamic programming approach to infinitely repeated games. That allows to obtain a set valued "value-iteration like" algorithm to compute the set of normalized equilibrium payoffs. This theory can be easily extended to stationary discounted dynamic games.

A second line of work, is provided in Harris [1985a] and [1985b]. In those papers, he studies the various topologies, other than the one introduced by Fudenberg and Levine, that may be used to obtain the set convergence result. Borgers[1989] took a different approach, by focusing on the space of feasible histories (i.e infinite sequences of outcomes). Assuming compactness of such space, he shows that the limit of a converging sequence of finite horizon equilibrium histories is an equilibrium history of the infinite horizon game.

Our objective in this paper, is to provide a new method to overcome "end of horizon" effects in the context of non-stationary infinite horizon dynamic games. Our inspiration comes from the works by Schochetman and Smith [1989] and [1992] which leads to the computational procedures described in [1991] and [1992]. The idea here is to restrict the deviation possibilities for players by forcing and ending target state. Interestingly enough, Ostrom, Gardner and Walker [1995] report the existence of a similar mechanism in the common pool resources exploitation. In that context, players agree to restrict the catch size or net size so that implicitly a reasonable level of stock is maintained.
On the other hand, recent work has pointed out the difficulties in extending to the dynamic game setting the sustainability of first best outcome as equilibrium play(see Dutta[1995]). Only under very restrictive assumptions (see Tolwinski[1982] and Gaitsgory and Nitzan[1994]) efficient outcomes have been shown to be sustained by Markovian Perfect Equilibrium. In addition to this, Engwerda[1996] has pointed out the various difficulties in approximating infinite horizon open-loop Nash equilibrium in Linear Quadratic games.
Motivated by all these features, we obtain a sequential characterization of Markov Perfect Infinite horizon equilibrium histories under some hypotheses. A tie-breaking rule is then proposed, as in Schochetman and Smith [1991] and [1992], to ensure convergence of a particular selection from the set of finite horizon relaxed solutions to a Markov Perfect Infinite horizon equilibrium history.

Although, the game structure assumed is more restrictive than that assumed by Fudenberg and Levine[1983] and Borgers[1989], the implementation of the computational procedure is simpler and it is extended to the undiscounted case, which constitutes a very significant advance.

This is the structure of the chapter. We introduce the class of nonstationary dynamic games that we study, then the new solution concept is presented and the results. Finally, an example for which the theory holds is presented. Large scale tests and computational implementation are currently being undertaken.

2 Nonstationary Dynamic Games.

For ease of the exposition, we will present the case for two players, bearing in mind that the extension to the Multi-player case is straightforward.

A two-person , discrete time, finite horizon dynamic game with state variable consists of :

- An index set $\mathcal{N} = \{0, 1, 2, 3....T\}$ denoting the stages of the game, where T is the maximum number of stages to be played, also known as "planning" horizon.
- An index set $\mathcal{I} = \{1, 2\}$ called the players set.
- A collection of sets $\{U_k^i : k \in \mathcal{N}\ i \in \mathcal{I}\}$, with $U_k^i \subset R^p$ for some p . We refer to this set as the "set of available controls at period k for player i"
- A collection of sets $\{S_k : k \in \mathcal{N}\}$, $S_k \subset R^q$ for some q, called the "set of attainable states at period k". The cartesian product $\prod_{k=0}^{T} S_k$ contains the set of T-long feasible sequences of states.
- A collection of functions $\{f_k : k \in \mathcal{N}\}$ of the form $f_k : S_k \times U_k^1 \times U_k^2 \mapsto S_{k+1}$, which models state dynamics :

$$s_{k+1} = f_k(s_k, u_k^1, u_k^2) \text{ for every } k \in \mathcal{N}$$

where s_0 is given a priori.

- Let $H(T)$ be defined as :

$$H(T) = \prod_{k=0}^{T} (U_k^1 \times U_k^2)$$

  In words, it is the set of T-long feasible sequences of controls that players exert. For every, $h_T \in H(T)$,one can recursively construct an element of $\prod_{k=0}^{T} S_k$ with the state dynamics presented above.
- A collection of functions $\{r_k^i : k \in \mathcal{N}\ i \in \mathcal{I}\}$ of the form $r_k^i : S_k \times U_k^1 \times U_k^2 \mapsto R$, called the "$k-$stage reward function for player i" .

- Discounted rewards : The total discounted reward for player $i \in I$ is :

$$P_T^i(h_T(s_0)) = \sum_{k=0}^{T} \lambda_i^k \cdot r_k^i(s_k, u_k^1, u_k^2)$$

where $\lambda_i \in [0, 1)$ is player's i discount factor.

- A closed loop perfect information structure, i.e at each time period players observe the "state" variable.

2.1 Strategies and Markov Perfect Equilibria (MPE).

We now introduce the concept of Nash Equilibria in strategies that employ available information for dynamic games with fixed "planning" horizon T .

- A *Closed-Loop Strategy* for player $i \in \mathcal{I}$, say π_i^T, is a $T-$tuple of continuous maps $\pi_k^i : S_{k-1} \rightarrow U_k^i$, so that π_i^T is of the form :

$$\pi_i^T = (\pi_0^i, \pi_1^i, ..., \pi_T^i)$$

We denote $\Pi^i(T)$ the set of all such strategies for player $i \in \mathcal{I}$. We refer to the 2-tuple $\pi^T = (\pi_1^T, \pi_2^T) \in \Pi^1(T) \times \Pi^2(T)$ as a *Closed Loop Strategy combination* and denote $\Pi(T)$ the set of all strategy combinations.

Such strategy combinations induce $T-$long feasible sequence of controls as follows :

1. In the first period players play $(\pi_0^1(s_0), \pi_0^2(s_0)) \in U_0^1 \times U_0^2$ so that at the end of that period the state attained is given by :

$$s_1 = f_0(s_0, \pi_0^1(s_0), \pi_0^2(s_0))$$

2. Then they observe $s_1 \in S_1$ and play $(\pi_1^1(s_1), \pi_1^2(s_1)) \in U_1^1 \times U_1^2$ so that the state to be attained is :

$$s_2 = f_2(s_1, \pi_1^1(s_1), \pi_1^2(s_1))$$

The process then unveils recursively.

We will denote by $h_T^{\pi^T}(s_0) \in H(T)$ the control sequence recursively obtained as above (sometimes called the "history" induced by strategy combination π^T) .
Note that from any state $s_k \in S_k$ with $k \neq 0$ and $k \in \mathcal{N}$,the strategy combination completely specifies the play that follows after state s_k recursively :

1. In period k players play $(\pi_k^1(s_k), \pi_k^2(s_k)) \in U_k^1 \times U_k^2$ so that at the end of that period the state attained is given by :

$$s_{k+1} = f_k(s_k, \pi_k^1(s_k), \pi_k^2(s_k))$$

2. Then they observe $s_k \in S_k$ and play $(\pi_{k+1}^1(s_k), \pi_{k+1}^2(s_k)) \in U_{k+1}^1 \times U_{k+1}^2$ so that the state to be attained is :

$$s_{k+2} = f_{k+1}(s_{k+1}, \pi_{k+1}^1(s_k), \pi_{k+1}^2(s_k))$$

The process then unveils recursively.

We will denote by $h_T^{\pi^T}(s_k) \in \prod_k^T (U_k^1 \times U_k^2)$ the control sequence recursively obtained as above (sometimes called the "history" induced by strategy combination π^T after state s_k) .

Remark 1 *We can conclude that a strategy combination can be alternatively described if for every attainable state $s_k \in S_k$ the play to follow is specified.*

We are now ready to introduce the solution concepts with which we shall be dealing with.

Definition 1 *We say that π^T is a* **Nash Equilibrium in Closed-Loop strategies** *iff for every player $i \in \mathcal{I}$ who would like to deviate from π^T by playing $\gamma_i^T \in \Pi^i(T)$ would find no incentive in doing so, i.e:*

$$P_T^i(h_T^{(\gamma_i^T, \pi_{-i}^T)}(s_0)) \leq P_T^i(h_T^{\pi^T}(s_0))$$

where $(\gamma_i^T, \pi_{-i}^T) \in \Pi(T)$ stands for the strategy combination in which all players $j \in \mathcal{I}$ and $j \neq i$ follow π_j^T and player $i \in \mathcal{I}$ follows γ_i^T .

The above solution concept is known not to be **time consistent** in the sense that the play prescribed after some state other than the initial state, may not constitute itself an equilibrium for the game that starts at such state. Hence, play off the equilibrium history is not "*credible*", that is, implicit "threats" of such off equilibrium play will not be taken seriously by opponents.To rule out such "non-credible threats" for deviators, a refinement of the previous solution concept is to require that the play to follow after any other state $s_k \in S_k$ with $0 \leq k < T$ must prescribe a Nash equilibrium in the sense above defined for the game that starts at s_k, most commonly known as a **subgame**(see Fudenberg and Tirole[1993]).

Definition 2 *We say that* π^T *is a* **Markov Perfect Equilibrium (MPE) in Closed-Loop strategies** *iff for every player* $i \in \mathcal{I}$ *who would like to deviate from* π^T *by playing* $\gamma_i^T \in \Pi^i(T)$ *from every* $s_k \in S_k$ *with* $0 \leq k < T$ *,would find no incentive in doing so, i.e:*

$$P_T^i(h_T^{(\gamma_i^T, \pi_{-i}^T)}(s_k)) \leq P_T^i(h_T^{\pi^T}(s_k))$$

where $(\gamma_i^T, \pi_{-i}^T) \in \Pi(T)$ *stands for the strategy combination in which all players* $j \in \mathcal{I}$ *and* $j \neq i$ *follow* π_j^T *and player* $i \in \mathcal{I}$ *follows* γ_i^T *.*

We denote $\Pi^*(T)$ the set of all "**Markov Perfect Nash equilibrium**" strategies.

Remark 2 *In the above definition we have slightly abused notation, by writing* $P_T^i(h_T^{(\pi_{-i}^T, \gamma_i^T)}(s_k))$ *since the play that follows after* s_k *, that is* $h_T^{(\pi_{-i}^T, \gamma_i^T)}(s_k)$ *is only defined in* $\prod_k^T (U_k^1 \times U_k^2)$ *. However, by the additive and separable reward structure assumed, the sense of the inequalities in the above definition is to be understood by appending "play" that reaches state* s_k *to* $h_T^{(\pi_{-i}^T, \gamma_i^T)}(s_k)$ *. It is worth pointing out that this is not correct in a more general setting, for instance, when there is no additive and separable reward structure assumed.*

2.2 Infinite Horizon Dynamic Games.

We now extend the above definitions and concepts to the case when $\mathcal{N} = \{0, 1, 2, ..\}$, i.e there is an infinite number of stages to play.

- A *Closed-Loop Strategy* for player $i \in I$, say π_i, is of the form $(\pi_0^i, \pi_1^i, ..., \pi_T^i, \pi_{T+1}^i, ...)$, where $\pi_k^i : S_{k-1} \rightarrow U_k^i$. We denote Π^i the set of all such strategies for player i.
  We refer to the 2-tuple $\pi = (\pi_1, \pi_2)$ as a *Closed Loop Strategy combination*, and denote Π the set of all closed loop strategy combinations.

As above explained, it should be clear that such combinations induce an infinite sequence of controls of the form $(\pi^0(s_0), \pi^1(s_1), ...) \in \prod_{k=0}^{\infty} (U_k^1 \times U_k^2) = H$. We will denote by $h^\pi(s_0)$ such sequence. Similarly, from any other state $s_k \in S_k$, the play that follows according to π is denoted by $h^\pi(s_k)$.

Moreover, the total aggregated reward received by player i under strategy combination π is defined by introducing the collections of maps $p^T : \Pi \rightarrow \Pi(T)$ which are the $T-$Horizon truncation of infinite horizon strategy combinations .:

$$P^i(h^\pi(s_0)) = \liminf_{T \rightarrow \infty} P_T^i(h_T^{p^T \pi}(s_0))$$

where $p^T\pi$stands for the $T-$period truncation of π .

All the above definitions carry over in a straigthforward manner.

2.3 Topologies on the set Π .

Since our interest is to study convergence of finite horizon equilibrium strategies to infinite horizon equilibrium strategies, it is very important to carefully define relevant topologies on Π, and consequently the different notions of convergence they induce.In this section we closely follow Harris[1985b] .

We will adopt the convention that any finite horizon strategy combination is trivially extended through any feasible choice of continuation sequence of strategies, so that its extension is an element of Π.

We first concentrate on a topology for H . Given $h = (u_0, u_1, u_2, ...,)$ and $h' = (u'_0, u'_1, u'_2, ...,)$ we define the metric $D : H \times H \to R^+$ by :

$$D(h, h') = \sup_t [\frac{\min\{d_t(u_t, u'_t), 1\}}{t}]$$

where $d_t : (U_t^1 \times U_t^2) \times (U_t^1 \times U_t^2) \to R$ is the product metric on $U_t^1 \times U_t^2$. This metric induces the product topology on H (see Munkres[1975], p. 123) .

Definition 3 *$\mathcal{W}$ is the topology with basis consisting of the sets :*

$$\{\pi \in \Pi \mid D(h^\pi, h) < \varepsilon\}$$

where $h \in H$, i.e some infinite horizon history. The basis is then obtained as we vary ε, h varies over H .

In words, the notion of convergence related to the topology $\mathcal{W}$ is simply the fact that for any given subgame (or state s_t) , $\pi^T \to \gamma$ with respect to $\mathcal{W}$ if and only if the sequence of histories induced by π^T, namely $\{h_T^{\pi^T}(s_t)\}_T$, converges to $h^\gamma(s_t)$ in the product topology.

Definition 4 *$\mathcal{L}$ is the topology with basis consisting of the sets :*

$$\{\pi \in \Pi \mid p^T\pi = p^T\gamma\}$$

with $\gamma \in \Pi$, obtained as γ varies over Π and T varies over all periods.

In words, $\pi^T \rightarrow \gamma$ with respect to $\mathcal{L}$ if and only if for all subgames simultaneously the sequence of histories induced by π^T converge in the discrete topology (they fully agree) to the histories induced by γ. $\mathcal{L}$ is essentially a uniform version of $\mathcal{W}$, for convergence in $\mathcal{L}$ requires that for all t there exists a $T(t)$ such that for all $T \geq T(t)$ we have that $p^t\pi^T = p^t\gamma$, that is, the strategy combinations from the first period up to period t as prescribed by π^T and γ agree. Clearly, for practical purposes it may be easier to prove convergence with respect to $\mathcal{W}$, since convergence need only be verified for representative subgames. On the contrary , $\mathcal{L}$ imposes more restrictive conditions on an approximating sequence, so it may be more helpful in proving uniqueness.

We now proceed to present the metric on Π that was used by Fudenberg and Levine[1983] and that for most applications will induce topology $\mathcal{L}$ on Π .
To capture the notion of closeness most relevant to Markovian Perfect Equilibrium we expect two strategies to be close if for every t and from every feasible state s_t (subgame) the histories induced by these strategies are close *and* the histories resulting when any one player deviates from them are also close. The metric that generates such topology is defined as follows, for $\pi, \gamma \in \Pi$:

$$\rho(\pi,\gamma) = \sup_{t, s_t \in S_t} \left\{ D(h^{\pi}(s_t), h^{\gamma}(s_t));\ \sup_{i, \delta_i \in \Pi_i} [D(h^{\delta_i, \pi_{-i}}(s_t), h^{\delta_i, \gamma_{-i}}(s_t))] \right\}$$

Fudenberg and Levine [1983] observed that the topology induced by metric ρ coincides with $\mathcal{L}$ when the action sets for all players in all periods are finite .

3 Constrained Markov Perfect Equilibria.

In this section, we introduce a new methodology to obtain sequential characterization of infinite horizon perfect equilibria.

3.1 Preliminaries.

We denote Π^*, Π^*_{st} and $\Pi^*(T)$: the set of Infinite Horizon MPE strategy combinations, the set of **strict** Infinite Horizon MPE strategy combinations and the set of $T-$period horizon MPE strategy combinations, respectively.

Definition 5 *Let $s_T \in S_T$ be some feasible state at time period T, we denote by $\Pi(T, s_T)$, the set of closed-loop strategy combination such that for every $s_k \in S_k$, $0 \leq k < T$, and the state s_T is reachable from s_k, the play to follow must reach s_T. In other words, the history prescribed*

.i.e $h_T^{\pi^T}(s_k)$ *reaches state* s_T, *at time period* T, *whenever state* s_T *is reachable from* $s_k \in S_k$, $0 \leq k < T$.

Note that the play prescribed by any $\pi^T \in \Pi(T, s_T)$ from some state $s_k \in S_k$ from which s_T is not reachable, is completely irrelevant to the definition.
Let us now briefly discuss the motivations for the solution concept relaxation that we will introduce shortly.
The main difficulty for a sequential characterization of infinite horizon equilibria as limits of finite horizon equilibria is due to "end of horizon" effects. In words, for a fixed finite horizon, the final state attained for finite horizon equilibrium will generally be different from the state attained by the truncation of the infinite horizon equilibrium . Myopic behavior close to the fixed finite horizon is the explanation for this.We will try to overcome this effect by forcing equilibrium strategies to attain a certain "target" state. However, this artificially imposed constraint introduces "strategic" alterations to the original game. Intuitively, some player may find it attractive to deviate in the early stages if he knows in advance that players will have to coordinate at the final stages in order to attain the target state. This is clearly a new artificial feature that as we will see pose difficulties to prove that any Infinite Horizon MPE is the limit of Finite Horizon Constrained Approximate MPE.

Definition 6 *A strategy combination* $\pi^T \in \Pi(T, s_T)$ *is called a* "**Constrained MPE to state** s_T" *iff for every deviation* $\gamma_i^T \in \Pi(T)$ *such that* $(\gamma_i^T, \pi_{-i}^T) \in \Pi(T, s_T)$ *from every* $s_k \in S_k$ *with* $0 \leq k < T$, *such that state* s_T *is reachable from* s_k *we have :*

$$P_T^i(h_T^{(\gamma_i^T, \pi_{-i}^T)}(s_k)) \leq P_T^i(h_T^{\pi^T}(s_k))$$

We denote $\Pi^*(T, s_T)$ the set of all "**Constrained MPE to state** s_T"

We now introduce the last additional notation needed for the analysis.

- Let :

$$\Lambda^*(T) = \bigcup_{s_T \in S_T} \Pi^*(T, s_T)$$

The set of all constrained MPE equilibrium strategies to all attainable states for horizon T.

4 A First Example : Sequential Duopoly.

In this section we briefly illustrate all the definitions above introduced for the case of a duopoly competition in prices, as in Maskin and Tirole [1988].
Players move sequentially, so that in odd numbered periods k , firm 1 chooses its price which remains unchanged until period $k+2$ .That is, $p^1_{k+1} = p^1_k$ if k is odd. Similarly, firm 2 chooses prices only in even numbered periods, $p^2_{k+1} = p^2_k$ if k is even. Hence, at time period k , firm's i instantaneous reward $r^i_k(.)$ is a function of the "state", i.e the price that firm's j set on period $k-1$, say p^j_k , and the "action" , i.e the price that firm's i will establish p^i_k . Price sets are discrete and bounded, goods are perfect substitutes, that is, firms share the market equally whenever they charge the same price. Firms have the same unit cost c. Let $D_k(.)$ denote the market demand function at time period k. The total reward at time period k is given by :

$$r_k(p) = (p-c)D_k(p)$$

Then :

$$r^i_k(p^1_k, p^2_k) = \begin{cases} r_k(p^i_k) & \text{if } p^i_k < p^j_k \\ \frac{r_k(p^i_k)}{2} & \text{if } p^i_k = p^j_k \\ 0 & \text{if } p^i_k > p^j_k \end{cases}$$

Strategies are "Markovian" in that they depend on the current "state", i.e last period rival's action. Hence, the set of all histories is the same as the set of all feasible sequences of states. Consider the infinite history $h = \{(p^1_k, p^2_k)\}_k$ then firm i total discounted payoff is :

$$P^i(h) = \sum_{k=0}^{\infty} \lambda^k_i \cdot r^i_k(p^1_k, p^2_k)$$

Now, let us assume that p^1_T is a feasible price decision for firm 1 at odd time period T. Then, $\Pi(T, p^1_T)$ stands for the set of all markovian strategy combinations for horizon T in which player 1 is constrained to play p^1_T at time period T .Similarly, $\Pi^*(T, p^1_T)$ is the set of T-long horizon "constrained" MPE strategy combinations to "state" p^1_T . Notice that under the assumptions by a backwards induction argument one can show that $\Pi^*(T, p^1_T) \neq \emptyset$.

5 A Second Example: Linear Quadratic Games.

Each player chooses controls at time period k, $u^i_k \in U^i_k \subset R^q$ where $i \in \mathcal{I}$, q some positive integer. The state variable $s_k \in S_k \subset R^p$ with p some positive integer, follows linear dynamics

$$s_k = As_{k-1} + \sum_{i \in \mathcal{I}} B_i u_k^i$$

where A is a given $p \times p$ matrix , and B_i $i \in \mathcal{I}$ are $p \times q$ given matrices. The initial value of the state vector s_0 is given.
Players payoffs are of the form :

$$r_k^i(s_k, u_k^i) = s_k' Q_k s_k + u_k' P_k u_k$$

where Q_k and P_k are $p \times p$ and $q \times q$ positive definite matrices respectively. Basar and Olsder [1995] give a wide array of applications of this model and sufficient conditions for existence of constrained open-loop Nash Equilibria.

6 Approximating Nonstationary MPE.

6.1 Assumptions.

Assumption 1: **(Non-Emptyness)** $\chi^*(T) \neq \emptyset$ for every T .

Assumption 2 :(Continuity) Reward functions are continuous and uniformly bounded, that is :

$$\forall\, k \in \mathcal{N}, i \in \mathcal{I} \quad s_k \in S_k, u_k \in U_k \quad r_k^i(s_k, u_k) \leq M < \infty$$

Note that $P_T^i \rightarrow P$ uniformly, so that $P(\cdot)$ is continuous.

Assumption 3: Reachability:
For any $s_k' \in S_k$, and any infinite feasible sequence of states $\mathbf{s} = (s_0, s_1, s_2, ...)$ with $s_k' \neq s_k$,there exist some finite time period $T > k$ and sequence of control profiles $\{u_s\}_{k<s\leq T}$, so that state s_T in the sequence is reached in $T = t + \triangle(s_k', \mathbf{s}) > t$ periods.

6.2 The Approximation Scheme.

We now present the main results in this section. Lemma 1(Harris[1985b]) ensures that in the family of dynamic games considered to prove that a certain strategy combination is an MPE for the Infinite Horizon game we only need to look at *finite* deviations, that is to say, deviations that take place for a finite number of periods. This result simplifies the task of proving that a certain candidate is in fact an MPE. On the other hand, Lemma 2 ensures that a limit point of the constrained MPE set constitutes a feasible strategy for the infinite horizon game. In words, the *reachability* assumption ensures that the limit point strategy is well defined.

6.2.1 Preliminaries.

Lemma 1 *Let $\pi \in \Pi$ and let us assume that no single period deviation from any subgame against π is profitable, then $\pi \in \Pi^*$, i.e it is a Markov Perfect Equilibrium.*

Proof. (Harris[1985b])Let $\gamma_i = (\gamma_0^i, \gamma_1^i, .., \gamma_T^i, \gamma_{T+1}^i, ...) \in \Pi^i$ be some general deviation for player i from π from the initial state s_0 .Let us construct a collection of deviations denoted by γ_i^k for player i that will approximate γ_i as follows.

$$\gamma_i^k = (\gamma_0^i, \gamma_1^i, .., \gamma_k^i, \pi_{k+1}^i, \pi_{k+2}^i ...)$$

By hypothesis :

$$P^i(h^{(\gamma_i^0, \pi_{-i})}(s_0)) \leq P^i(h^\pi(s_0))$$

Let us denote by s_k the state attained by following deviation γ_i^k from π with $0 \leq k \leq T$ .Then by hypothesis :

$$P^i(h^{(\gamma_i^{k+1}, \pi_{-i})}(s_k)) \leq P^i(h^\pi(s_k))$$

Concatenating these inequalities we obtain :

$$P^i(h^{(\gamma_i^T, \pi_{-i})}(s_0)) \leq P^i(h^\pi(s_0))$$

Then by construction :

$$h^{(\gamma_i^T, \pi_{-i})}(s_0) \rightarrow h^{(\gamma_i, \pi_{-i})}(s_0) \quad \text{as } T \rightarrow \infty$$

and continuity of discounted reward functional :

$$P^i(h^{(\gamma_i, \pi_{-i})}(s_0)) \leq P^i(h^\pi(s_0))$$

The same argument can be argued if the deviation "starts" at an arbitrary intermediate feasible state .

■

The idea behind the above presented proof is simple. If single period deviations are unprofitable then finite sequences of deviations unravel. Since any infinite sequence of deviations gives rise to a history that is the limit of histories derived from finite sequences of deviations. None of these deviations are unprofitable and the limit is also not profitable by continuity.
The next lemma ensures that a limit point from our constrained MPE set is a well defined strategy for the Infinite Horizon game.

Lemma 2 $\limsup_{T\to\infty} \chi^*(T) \subset \Pi$.

Proof. Let $\pi \in \limsup_{T\to\infty} \chi^*(T)$. By definition, there is a infinite feasible sequence of states

$$\{s_T : T \in \mathcal{N}\}$$

and constrained MPE strategy combinations $\{\pi^T : T \in \mathcal{N}\}$ with $\pi^T \in \Pi^*(T, s_T)$, such that the it has a converging subsequence $\{\pi^{T_k} : k \in \mathcal{K} \subset \mathcal{N}\}$ with respect to the topology $\mathcal{W}$ whose limit is π.
We will focus our attention on the following infinite sequence of states :

$$\mathsf{s} = (..., s_{T_k}, s'_{T_k+1} = f_{T_k}(s_{T_k}, \pi_{T_{k+1}}(s_{T_k})), s'_{T_k+2} = f_{T_k+1}(s'_{T_k+1}, \pi_{T_{k+1}}(s'_{T_k+1}), ..., s_{T_{k+1}}, ...,)$$

In words, we have appropiately filled the gaps on the sequence $\{s_{T_k}\}_{k\in\mathcal{K}}$.

In order to prove that $\pi \in \Pi$, we need to verify that from an arbitrary intermediate feasible state s_m, the play prescribed by π is well defined.
By **reachability** from state s_m there exist a feasible finite sequence of actions the "reaches" the infinite sequence of states s above constructed at some period n . Then let :

$$k^* = \min\{k \in \mathcal{K} : T_k \geq n\}$$

Thus every π^{T_k} such that $k \geq k^*$ prescribes play that reaches state s_{T_k} which we denote by :

$$h_{T_k}^{\pi^{T_k}}(s_m)$$

By definition of convergence in $\mathcal{W}$:

$$h_{T_k}^{\pi^{T_k}}(s_m) \to h^{\pi}(s_m)$$

So we conclude that $\pi \in \Pi$. ■

"The limit point of constrained MPE is an MPE for the Infinite Horizon Game".

Theorem 3 $\limsup_{T\to\infty} \chi^*(T) \subset \Pi^*$ *with respect to the topology* $\mathcal{W}$.

Proof.

Let us first show that :

$$P^i(h^{(\pi_{-i},\gamma_i)}(s_k)) \leq P^i(h^{\pi}(s_k))$$

for any player $i \in \mathcal{I}$, who would deviate by playing $\gamma_i \in \Pi^i$, which constitutes a single period deviation from π from the arbitrary feasible state s_m . We recall that this is sufficient in view of Lemma 1

Then by convergence in $\mathcal{W}$ and lemma 2 :

$$h_{T_k}^{(\gamma_i^{T_k},\pi_{-i}^{T_k})}(s_m) \rightarrow h^{(\gamma_i,\pi_{-i})}(s_m) \text{ as } k \rightarrow \infty$$

where $(\gamma_i^{T_k}, \pi_{-i}^{T_k})$ is simply the strategy combination formed of the T_k-period truncation of γ_i as a deviation from π^{T_k} .

.By hypothesis :

$$P_{T_k}^i(h_{T_k}^{(\gamma_i^{T_k},\pi_{-i}^{T_k})}(s_m)) \leq P_{T_k}^i(h_{T_k}^{\pi^{T_k}}(s_m))$$

Finally ,by continuity of the discounted reward functional after taking limits as $k \rightarrow \infty$:

$$P^i(h^{(\gamma_i,\pi_{-i})}(s_m)) \leq P^i(h^{\pi}(s_m))$$

■

We note that Lemma 2 and the above theorem are easily proven when considering the topology $\mathcal{L}$ in the set of Infinite Horizon strategies.

Corollary 4 *Lemma 2 and Theorem 1 hold true with respect to the stronger topology* $\mathcal{L}$.

"Any MPE is the limit of Constrained MPE" .

Let $p^T\pi$ be the $T-$truncation of π .Let s_T be the state attained by $h_T^{p^T\pi}(s_0)$, in general :

$$p^T\pi \notin \Pi(T, s_T)$$

The reason being, that from some state $s_k \in S_k$ with $0 < k < T$ the play prescribed by $p^T\pi$, i.e $h_T^{p^T\pi}(s_k)$ need not attain state s_T at time period T .We now define a projection operation which, when applied to any strategy combination π^T, will yield the "closest" constrained strategy combination to π^T.

Definition 7 *Let $\pi^T \in \Pi(T)$ be some strategy combination that attains state s_T at time period T .The projection $\mu_T(\pi^T) = \hat{\pi}^T$ of π^T is defined as follows :*

- The play prescribed by $\hat{\pi}^T$ from s_0 is exactly the one prescribed by π^T . That is :

$$h_T^{\hat{\pi}^T}(s_0) = h_T^{\pi^T}(s_0)$$

- For all states $s_k \in S_k$ with $0 < k < T$ such that $h_T^{\pi^T}(s_k)$ reaches state s_T at time period T, we again set $h_T^{\hat{\pi}^T}(s_k)$ to be exactly $h_T^{\pi^T}(s_k)$.

- For all states $s_k \in S_k$ with $0 < k < T$ such that $\pi_k^T(s_k)$ leads to state s_{k+1} (through, the dynamics $s_{k+1} = f_k(s_k, \pi_k^T(s_k))$) from where state s_T is reachable, we set

$$\hat{\pi}_k^T(s_k) = \pi_k^T(s_k)$$

- For any other state $s_k \in S_k$ with $0 < k < T$ not in any of the above cases, and such that state s_T is reachable from s_k we set $h_T^{\hat{\pi}^T}(s_k)$ to be **any sequence of controls required to attain** state s_T.

Notice that by construction $\hat{\pi}^T \in \Pi(T, s_T)$. Intuitively, this is the constrained strategy combination to state s_T whose play from every feasible state s_t $t \leq T$ "**resembles**" the most the play prescribed by π^T. However, as we shall now see, the great degree of freedom in choosing ending play that satisfies the ending target state requirement renders difficult the full sequential characterization.

We now prove the converse statement.

Theorem 5 *If Π is compact with respect to $\mathcal{W}$ then*

$$\Pi_{st}^* \subset \lim \inf_{T \to \infty} \chi^*(T)$$

Proof. Let π^T be the $T-$truncation of $\pi \in \Pi_{st}^*$ and s_T the state attained by following this strategy.Let us consider the sequence $\{\hat{\pi}^T : T \in \mathcal{N}\}$ where $\hat{\pi}^T = \mu_T(\pi^T)$.

By construction we have :

$$\pi = \lim_{T \to \infty} \hat{\pi}^T$$

Reasoning by contradiction, let us now assume that :

$$\hat{\pi}^T \notin \Pi^*(T, s_T) \text{ for all } T$$

By definition, this implies the existence of profitable deviations, i.e for each T there is γ_i^T with $i \in \mathcal{I}$ such that $(\gamma_i^T, \hat{\pi}_{-i}^T) \in \Pi(T, s_T)$ and :

$$P_T^i(h_T^{(\gamma_i^T, \hat{\pi}_{-i}^T)}(s_0)) > P_T^i(h_T^{\hat{\pi}^T}(s_0))$$

But by compactness of the space Π with respect to the topology $\mathcal{W}$, the collection :

$$\{(\gamma_i^T, \hat{\pi}_{-i}^T) : T \subset \mathcal{N} \quad i \in \mathcal{I}\}$$

has a converging subsequence with limit, say $(\gamma_i, \pi_{-i}) \in \Pi$. By continuity of discounted reward functional after taking limits, we obtain :

$$P^i(h^{(\gamma_i, \pi_{-i})}(s_0)) \geq P^i(h^{\pi}(s_0))$$

In other words, $\pi \notin \Pi_{st}^*$, hence a contradiction. ■

A comment on this result is mandatory: we have said that the artificially imposed constraint of reaching a target state, introduced "strategic" alterations to the original game. Intuitively, some player may find it attractive to deviate in the early stages if he knows in advance that players will have to coordinate at the final stages in order to attain the target state. Thus it is critical to choose when constructing the projection map $\mu_T(.)$ the right ending play. However, as the planning horizon diverges to infinity this alteration becomes negligible. Nonetheless, only "strict" Infinite Horizon Markov Perfect Equilibria are immune to this effect.

We believe that this problem can be solved by looking at the specifics of each application so to reduce the artificiality involved in constructing the projection map $\mu_T(.)$. This is a matter of further research.

Theorem 6 *If all Infinite Horizon Markov Perfect Equilibria are* **"strict"** *and* Π *is compact in* $\mathcal{W}$ *then :*

$$\lim_{T \to \infty} \chi^*(T) = \Pi^*$$

Proof. By the statement in the hypothesis and theorems 1 and 2 :

$$\limsup_{T \to \infty} \chi^*(T) \subset \Pi^* = \Pi_{st}^* \subset \liminf_{T \to \infty} \chi^*(T)$$

■

7 Existence of Markov Perfect Equilibria (Average Reward).

Perhaps the main contribution of the approach introduced is that it enables to infer existence in the undiscounted case, provided the set of constrained MPE for finite planning horizon is not empty. We say that it constitutes a significant advance in that none of the sequential characterizations available and reviewed, can be applied in this context.

We focus our attention for the case when players do not discount rewards, i.e $\lambda_i = 1$ for $i \in \mathcal{I}$. For the infinite horizon dynamic game we define the following total aggregated reward received by player i under strategy combination π as follows :

$$A^i(h^\pi(s_0)) = \lim_{T\to\infty} \inf \frac{P_T^i(h_T^{p^T\pi}(s_0))}{T}$$

Since there is no discounting, the function $P_T^i(.)$ yields the total reward accrued up to period T . Thus, the above defined functional can be interpreted as the "average reward" received under strategy combination π from initial state s_0. One can easily extend this definition if we restrict to different starting states at different periods.

7.1 Assumptions.

Assumption 1: **(Non-Emptyness)** $\chi^*(T) \neq \emptyset$ for every T .

Assumption 2 (Average case):

1. **Discreteness** . Each set U_k^i is discrete, and endowed with the discrete metric. Hence, the product space $U = \prod_{k=0}^{\infty} U_k^1 \times U_k^2$ is compact in the product topology and Π is compact in $\mathcal{L}$.
2. **Reward Boundedness** , that is, for every $i \in \mathcal{I}$ and $k \in \mathcal{N}$:

$$-\infty < -M \leq r_k^i(\cdot,\cdot) \leq M < \infty$$

Assumption 3' Uniformly Bounded Reachability:
For any $s_k' \in S_k$, and any infinite feasible sequence of states $\mathbf{s} = (s_0, s_1, s_2, ...)$ with $s_k' \neq s_k$,there

exist some finite time period $T > k$ and sequence of control profiles $\{u_s\}_{k<s\leq T}$, so that state s_T in the sequence is reached in $T = k + \triangle(s'_k, \mathbf{s}) > t$ periods. Moreover,

$$\sup_k \sup_{\mathbf{s}} \triangle(s'_k, \mathbf{s}) \leq L < \infty$$

A comment is due on this last assumption. In the discounted case, the reachability assumption required players to be able to effectively control the state dynamics by cooperating, i.e given any infinite feasible sequence of states and any state off that sequence, players could agree on a finite sequence of actions so to *reach* the given infinite sequence. In the average case, we have to limit the "reward" effect of such finite sequence of actions by requiring that it is of an uniformly bounded length. This assumption then ensures that in the average the "reward" effect caused by it disappears.

7.2 The Existence Result.

Theorem 7 $\emptyset \neq \limsup_{T\to\infty} \chi^*(T) \subset \Pi^*$ *with respect to* $\mathcal{L}$.

Proof. Non-emptyness is due to discreteness of action sets.

Let $\pi \in \limsup_{T\to\infty} \chi^*(T)$. By Lemma 2 , we know that $\pi \in \Pi$.
Let us denote by $\{\pi^k : k \in \mathcal{K} \subset \mathcal{N}\}$ such that :

$$\pi = \lim_{k\to\infty} \pi^k \text{ with respect to } \mathcal{L}$$

Let us first show that :

$$A^i(h^{(\gamma_i,\pi_{-i})}(s_0)) \leq A^i(h^\pi(s_0))$$

for any player $i \in \mathcal{I}$, who would deviate by playing $\gamma_i \in \Pi$ from initial state s_0. We recall that $h_T^{(\gamma_i,\pi_{-i})}(s_0)$ and $h_T^\pi(s_0)$ stand for the $T-$truncations of such histories.
Let s'_T denote the state attained by $h_T^{(\gamma_i,\pi_{-i})}(s_0)$.

By convergence in $\mathcal{L}$ and Lemma 2 there exist $k_T \in \mathcal{K}$ such that for any π^k with $k \geq k_T > T$the play prescribed by (γ_i^k, π_{-i}^k) and π^k coincide exactly with $h_T^{(\gamma_i,\pi_{-i})}(s_0)$ and $h_T^\pi(s_0)$ respectively, in the first T periods . Moreover, the deviation for player i :

$$\gamma_i^k = (\gamma_0^i, \gamma_1^i, ..., \gamma_T^i, a_{T+1}^i, ..., a_k^i)$$

in which we append from $T-$period the actions $(a^i_{T+1}, ..., a^i_{T+1})$ as prescribed π^k, is such that $(\bar{\gamma}^k_i, \pi^k_{-i})$ "**reaches**" state s_k. Formally :

$$(\bar{\gamma}^k_i, \pi^k_{-i}) \in \Pi(k, s_k)$$

Hence, by hypothesis on π^k with $k \geq k_T > T$ we have :

$$P^i_k(h_k^{(\bar{\gamma}^k_i, \pi^k_{-i})}(s_0)) \leq P^i_k(h_k^{\pi^k}(s_0))$$

By cost boundedness and the choice of k, the strict reachability assumption and the *fact that we deal with markovian strategies* :

$$\frac{P^i_T(h_T^{(\bar{\gamma}^k_i, \pi^k_{-i})}(s_0))}{T} \leq \frac{P^i_T(h_T^{\pi^k}(s_0))}{T} + \frac{2M \cdot L}{T}$$

and :

$$\frac{P^i_T(h_T^{(\gamma_i, \pi_{-i})}(s_0))}{T} \leq \frac{P^i_T(h_T^{\pi}(s_0))}{T} + \frac{2M \cdot L}{T}$$

Hence :

$$A^i(h^{(\gamma_i, \pi_{-i})}(s_0)) = \lim_{T\to\infty} \inf \frac{P^i_T(h_T^{(\gamma_i, \pi_{-i})}(s_0))}{T} \leq \lim_{T\to\infty} \inf \frac{P^i_T(h_T^{\pi}(s_0))}{T} = A^i(h^{\pi}(s_0))$$

Thus, from the initial state, the proposed deviation is not profitable.

For a deviation from any other state $s_k \in S_k$ with $0 < k$ we proceed in the same manner. ■

8 Conclusion.

In this chapter, we have addressed the problem of approximating MPE for Infinite Horizon Games. For finite planning horizons, we introduced the notion of "Constrained MPE" as a surrogate that is immune to "end of horizon" effects.
We see that under fairly general assumptions and "reachability" (players can, by cooperating, effectively control the state dynamics) the limit point of constrained MPE is an MPE for the Infinite Horizon Game. This result is of particular value as a computational procedure.
However, due to "strategic" effects induced by the artificial terminal constraint, any Infinite Horizon MPE is not necessarily the limit of "Constrained MPE". Nonetheless, when all Infinite

Horizon MPE are **strict** and the space of strategies is compact, then every Infinite Horizon MPE is the limit of "Constrained MPE". This sequential characterization allows for computational procedures as in Smith and Schochetman [1991] and [1992].
We believe that substantial improvements can be obtained by focusing in very specific settings in order to avoid the artificiality resulting from working in an abstract framework. This is a subject of further research.

Finally, and perhaps more importantly, for a different set of assumptions, when players do not use discounting, the limit point of "Constrained MPE" is MPE for the Infinite Horizon Game. A standard compactness argument yields existence, which constitutes a significant advance given that existing methodologies do not work well in this setting.

References

[1] "Infinite Horizon Optimization", Schochetman I. and Smith R.L. *Mathematics of Operations Research* 14 (1989) 1-16

[2] "Finite dimensional approximation in Infinite dimensional mathematical programming", Schochetman I. and Smith R. L.*Mathematical Programming* 54(1992) 307-333

[3] "Convergence of Selections with applications to Optimization", Schochetman I., Smith R.L.
J. Math . Anal. Appl. 155 (1991) 278-292

[4] "Convergence of Best Approximations from Unbounded Sets" Schochetman I., Smith R.L.
J. Math . Anal. Appl. 166 (1992) pp. 112-128

[5] "Subgame-Perfect equilibria of Finite and Infinite-Horizon Games", D.Fudenberg and D.Levine
Journal of Economic Theory 31 (1983) 251-267

[6] "Limit Games and Limit Equilibria", D.Fudenberg and D.Levine
Journal of Economic Theory 38 (1986) 261-279

[7] "Collusive Behavior in non-cooperative epsilon-equilibria of Oligopolies with long but finite lives"
Journal of Economic Theory 22 (1980) pp 136-154

[8] "Theory of Correspondances", E. Klein. *John Wiley* 1984

[9] "Topology : A first course", J.R Munkres. *Prentice Hall* 1975

[10] "Perfect Equilibrium histories of Finite and Infinite Horizon Games" Börgers,Tilman.
Journal of Economic Theory 47 (1989) pp 218-227

[11] "A characterization of perfect equilibria in Infinite Horizon Games" Harris,Chris.
Journal of Economic Theory 37 (1985) pp. 99-127

[12] "Existence and Characterization of Perfect equilibrium in Games of Perfect Information" Harris,Chris.
Econometrica 53 No 3 (1985) pp. 613-628

[13] "Cooperation and Rationality" D. Pearce
Sixth World Congress of the Econometrica Society (1990), Jean Jacques Laffont ed.

[14] "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring" Abreu, D., Pearce, D. and Stacchetti E.
Econometrica 58 No 5 (1990) pp. 1041-1063

[15] "Optimal Cartel Equilibria with Imperfect Monitoring" Abreu, D., Pearce, D. and Stacchetti E.
Journal of Economic Theory 39 (1986) pp. 251-269

[16] "A Theory of Dynamic Oligopoly II : Price Competition, Kinked Demand curves and Edgeworth cycles" Maskin E. and Tirole J.
Econometrica 56 No.3 (1988) 571-599

[17] "Game Theory" Fudenberg D. and Tirole J.
MIT Press 1993

[18] "Dynamic Non-Cooperative Game Theory" Basar T. and Olsder G.
Academic Press 1995 (2nd Edition)

[19] "A folk theorem for dynamic games" Gaitsgory V. and Nitzan Sh.
Journal of Mathematical Economics 23 (1994) pp 167-178

[20] "A concept of cooperative equilibrium for dynamic games" Tolwinsky B.
Automatica 18 (1982) pp. 431-446

[21] "Computational aspects of the Infinite planning horizon open-loop Nash equilibrium in LQ games" Engwerda J.C
Working paper 1996, *Tilburg University*