An Introduction to Dynamic Games

A. Haurie J. Krawczyk

March 28, 2000

Contents

Chapter 1

Foreword

1.1 What are Dynamic Games?

Dynamic Games are mathematical models of the interaction between different agents who are controlling a dynamical system. Such situations occur in many instances like armed conflicts (e.g. duel between a bomber and a jet fighter), economic competition (e.g. investments in R&D for computer companies), parlor games (Chess, Bridge). These examples concern dynamical systems since the actions of the agents (also called players) influence the evolution over time of the *state* of a system (position and velocity of aircraft, capital of know-how for Hi-Tech firms, positions of remaining pieces on a chess board, etc). The difficulty in deciding what should be the behavior of these agents stems from the fact that each *action* an agent takes at a given time will influence the reaction of the *opponent(s)* at later time. These notes are intended to present the basic concepts and models which have been proposed in the burgeoning literature on game theory for a representation of these dynamic interactions.

1.2 Origins of these Lecture Notes

These notes are based on several courses on **Dynamic Games** taught by the authors, in different universities or summer schools, to a variety of students in engineering, economics and management science. The notes use also some documents prepared in cooperation with other authors, in particular B. Tolwinski [Tolwinski, 1988].

These notes are written for **control engineers**, **economists** or **management scientists** interested in the analysis of multi-agent optimization problems, with a particular

emphasis on the modeling of conflict situations. This means that the level of mathematics involved in the presentation will not go beyond what is expected to be known by a student specializing in control engineering, quantitative economics or management science. These notes are aimed at last-year undergraduate, first year graduate students.

The Control engineers will certainly observe that we present *dynamic games* as an extension of *optimal control* whereas economists will see also that *dynamic games* are only a particular aspect of the *classical theory of games* which is considered to have been launched in [Von Neumann & Morgenstern 1944]. Economic models of imperfect competition, presented as variations on the "classic" Cournot model [Cournot, 1838], will serve recurrently as an illustration of the concepts introduced and of the theories developed. An interesting domain of application of *dynamic games*, which is described in these notes, relates to environmental management. The conflict situations occurring in fisheries exploitation by multiple agents or in policy coordination for achieving global environmental control (e.g. in the control of a possible global warming effect) are well captured in the realm of this theory.

The objects studied in this book will be *dynamic*. The term *dynamic* comes from Greek *dynasthai* (which means *to be able*) and refers to phenomena which undergo a time-evolution. In these notes, most of the dynamic models will be *discrete time*. This implies that, for the mathematical description of the dynamics, *difference* (rather than *differential*) equations will be used. That, in turn, should make a great part of the notes accessible, and attractive, to students who have not done advanced mathematics. However, there will still be some developments involving a *continuous time* description of the dynamics and which have been written for readers with a stronger mathematical background.

1.3 Motivation

There is no doubt that a course on dynamic games suitable for both control engineering students and economics or management science students requires a specialized textbook.

Since we emphasize the detailed description of the dynamics of some specific systems controlled by the players we have to present rather sophisticated mathematical notions, related to control theory. This presentation of the dynamics must be accompanied by an introduction to the specific mathematical concepts of game theory. The originality of our approach is in the mixing of these two branches of applied mathematics.

There are many good books on *classical* game theory. A nonexhaustive list in-

cludes [Owen, 1982], [Shubik, 1975a], [Shubik, 1975b], [Aumann, 1989], and more recently [Friedman 1986] and [Fudenberg & Tirole, 1991]. However, they do not introduce the reader to the most general *dynamic* games. [Başar & Olsder, 1982] does cover extensively the dynamic game paradigms, however, readers without a strong mathematical background will probably find that book difficult. This text is therefore a modest attempt to bridge the gap.

Part I

Elements of Classical Game Theory

Chapter 2

Decision Analysis with Many Agents

As we said in the introduction to these notes *dynamic games* constitute a subclass of the mathematical models studied in what is usually called the *classical theory of game*. It is therefore proper to start our exposition with those basic concepts of game theory which provide the fundamental tread of the theory of dynamic games. For an exhaustive treatment of most of the definitions of classical game theory see *e.g.* [Owen, 1982], [Shubik, 1975a], [Friedman 1986] and [Fudenberg & Tirole, 1991].

2.1 The Basic Concepts of Game Theory

In a *game* we deal with the following concepts

- *Players*. They will *compete* in the game. Notice that a player may be an individual, a set of individuals (or a *team* , a corporation, a political party, a nation, a pilot of an aircraft, a captain of a submarine, *etc.* .
- A *move* or a *decision* will be a player's action. Also, borrowing a term from control theory, a move will be realization of a player's control or, simply, his *control*.
- A player's (pure) *strategy* will be a rule (or function) that associates a player's move with the information available to him[1] at the time when he decides which move to choose.

[1] Political correctness promotes the usage of gender inclusive pronouns "they" and "their". However, in games, we will frequently have to address an individual player's action and distinguish it from a collective action taken by a set of several players. As far as we know, in English, this distinction is only possible through usage of the traditional grammar gender exclusive pronouns: possessive "his", "her" and personal "he", "she". We find that the traditional grammar better suits your purpose (to avoid)

- A player's *mixed strategy* is a probability measure on the player's space of pure strategies. In other words, a mixed strategy consists of a random draw of a pure strategy. The player controls the probabilities in this random experiment.
- A player's *behavioral strategy* is a rule which defines a random draw of the admissible move as a function of the information available[2]. These strategies are intimately linked with mixed strategies and it has been proved early [Kuhn, 1953] that, for many games the two concepts coincide.
- *Payoffs* are real numbers measuring desirability of the possible outcomes of the game, *e.g.* , the amounts of money the players may win (or loose). Other names of payoffs can be: *rewards*, *performance indices* or *criteria*, *utility measures*, *etc.* .

The concepts we have introduced above are described in relatively imprecise terms. A more rigorous definition can be given if we set the theory in the realm of *decision analysis* where *decision trees* give a representation of the dependence of *outcomes* on *actions* and *uncertainties* . This will be called the *extensive form of a game.*

2.2 Games in Extensive Form

A game in extensive form is a graph (*i.e.* a set of nodes and a set of arcs) which has the structure of a tree[3] and which represents the possible sequence of actions and random perturbations which influence the outcome of a game played by a set of players.

2.2.1 Description of moves, information and randomness

A game in extensive form is described by a set of players, including one particular player called *Nature* , and a set of *positions* described as *nodes* on a tree structure. At each node one particular player has the right to move, *i.e.* he has to select a possible action in an admissible set represented by the arcs emanating from the node.

The *information* at the disposal of each player at the nodes where he has to select an action is described by the *information structure of the game* . In general the player

confusion and we will refer in this book to a singular genderless agent as "he" and the agent's possession as "his".

[2] A similar concept has been introduced in control theory under the name of *relaxed controls.*

[3] A *tree* is a graph where all nodes are connected but there are no cycles. In a tree there is a single node without "parent", called the "root" and a set of nodes without descendants, the "leaves". There is always a single path from the root to any leaf.

may not know exactly at which node of the tree structure the game is currently located. His information has the following form:

> he knows that the current position of the game is an element in a given subset of nodes. He does not know which specific one it is.

When the player selects a move, this correponds to selecting an arc of the graph which defines a transition to a new node, where another player has to select his move, etc. Among the players, *Nature* is playing randomly, i.e. Nature's moves are selected at random. The game has a stopping rule described by terminal nodes of the tree. Then the players are paid their rewards, also called *payoffs* .

Figure 2.1 shows the extensive form of a two-player, one-stage stochastic game with simultaneous moves. We also say that this game has the *simultaneous move information structure* . It corresponds to a situation where Player 2 does not know which action has been selected by Player 1 and vice versa. In this figure the node marked D_1 corresponds to the move of player 1, the nodes marked D_2 correspond to the move of Player 2.

The information of the second player is represented by the oval box. Therefore Player 2 does not know what has been the action chosen by Player 1. The nodes marked E correspond to Nature's move. In that particular case we assume that three possible elementary events are equiprobable. The nodes represented by dark circles are the terminal nodes where the game stops and the payoffs are collected.

This representation of games is obviously inspired from *parlor games* like *Chess* , *Poker* , *Bridge* , etc which can be, at least theoretically, correctly described in this framework. In such a context, the randomness of *Nature* 's play is the representation of card or dice draws realized in the course of the game.

The extensive form provides indeed a very detailed description of the game. It is however rather non practical because the size of the tree becomes very quickly, even for simple games, absolutely huge. An attempt to provide a complete description of a complex game like *Bridge* , using an extensive form, would lead to a combinatorial explosion. Another drawback of the extensive form description is that the states (nodes) and actions (arcs) are essentially finite or enumerable. In many models we want to deal with, actions and states will also often be continuous variables. For such models, we will need a different method of problem description.

Nevertheless extensive form is useful in many ways. In particular it provides the fundamental illustration of the dynamic structure of a game. The ordering of the sequence of moves, highlighted by extensive form, is present in most games. *Dynamic games* theory is also about *sequencing* of actions and reactions. Here, however, dif-

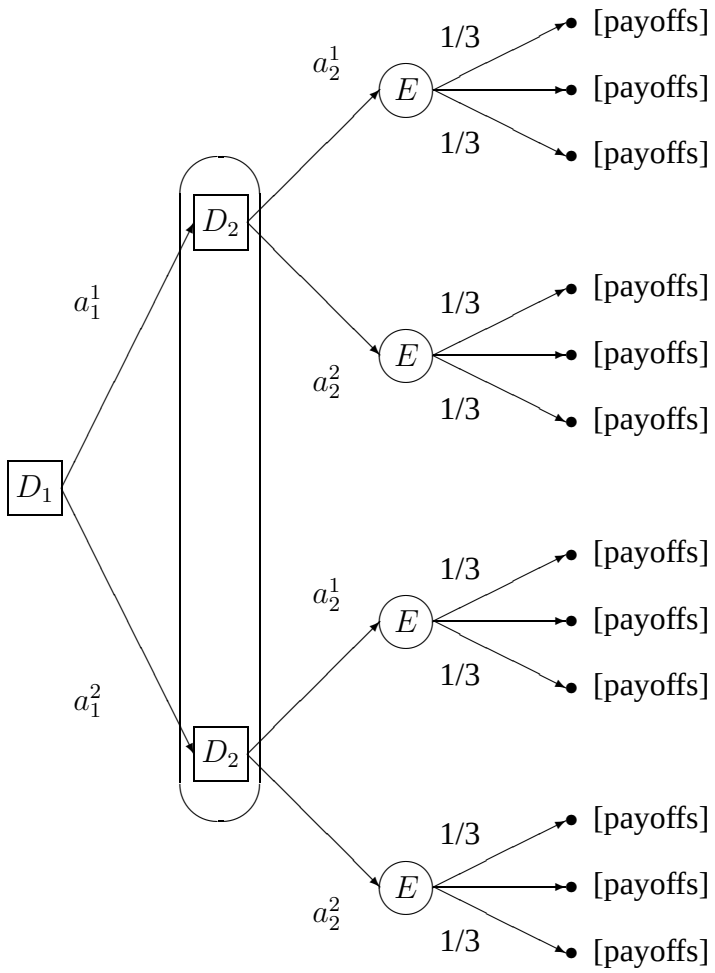


Figure 2.1: A game in extensive form

ferent mathematical tools are used for the representation of the game dynamics. In particular, differential and/or difference equations are utilized for this purpose.

2.2.2 Comparing Random Perspectives

Due to Nature's randomness, the players will have to compare and choose among different *random perspectives* in their decision making. The fundamental decision structure is described in Figure 2.2. If the player chooses action a_1 he faces a random perspective of expected value 100. If he chooses action a_2 he faces a sure gain of 100. If the player is *risk neutral* he will be indifferent between the two actions. If he is *risk*

averse he will choose action a_2, if he is *risk lover* he will choose action a_1. In order to

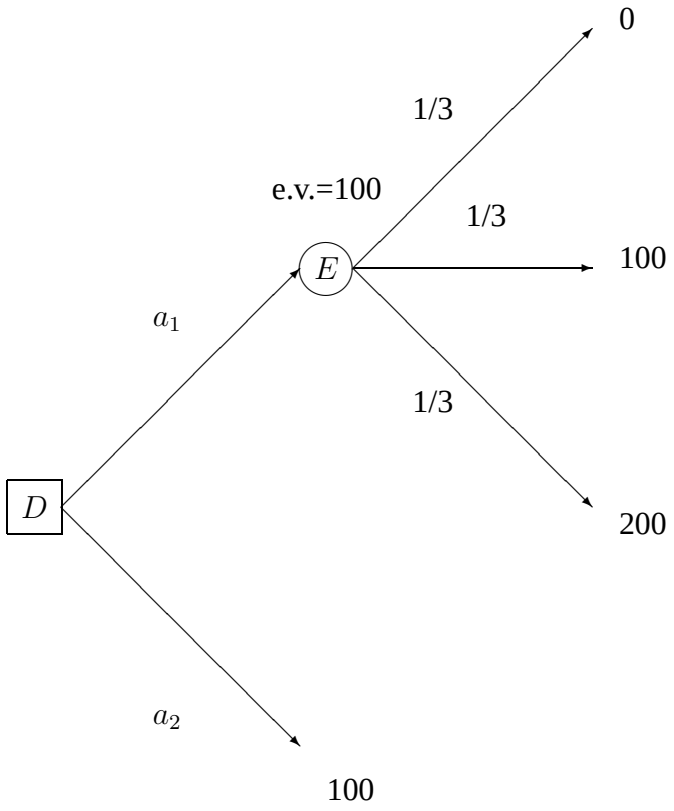


Figure 2.2: Decision in uncertainty

represent the attitude toward risk of a decision maker Von Neumann and Morgenstern introduced the concept of *cardinal utility* [Von Neumann & Morgenstern 1944]. If one accepts the axioms of utility theory then a *rational player* should take the action which leads toward the random perspective with the highest *expected utility* .

This solves the problem of comparing random perspectives. However this also introduces a new way to play the game. A player can set a random experiment in order to generate his decision. Since he uses utility functions the principle of maximization of expected utility permits him to compare deterministic action choices with random ones.

As a final reminder of the foundations of utility theory let's recall that the Von Neumann-Morgenstern utility function is defined up to an affine transformation. This says that the player choices will not be affected if the utilities are modified through an affine transformation.

2.3 Additional concepts about information

What is known by the players who interact in a game is of paramount importance. We refer briefly to the concepts of complete and perfect information.

2.3.1 Complete and perfect information

The information structure of a game indicates what is known by each player at the time the game starts and at each of his moves.

Complete vs Incomplete Information

Let us consider first the information available to the players when they enter a game play. A player has *complete information* if he knows

- who the players are
- the set of actions available to all players
- all possible outcomes to all players.

A game with *complete information* and *common knowledge* is a game where all players have complete information and all players know that the other players have complete information.

Perfect vs Imperfect Information

We consider now the information available to a player when he decides about specific move. In a game defined in its extensive form, if each information set consists of just one node, then we say that the players have *perfect information* . If that is not the case the game is one of *imperfect information* .

Example 2.3.1 *A game with simultaneous moves, as e.g. the one shown in Figure 2.1, is of imperfect information.*

Perfect recall

If the information structure is such that a player can always remember all past moves he has selected, and the information he has received, then the game is one of *perfect recall* . Otherwise it is one of *imperfect recall* .

2.3.2 Commitment

A *commitment* is an action taken by a player that is binding on him and that is known to the other players. In making a commitment a player can persuade the other players to take actions that are favorable to him. To be effective commitments have to be *credible* . A particular class of commitments are *threats* .

2.3.3 Binding agreement

Binding agreements are restrictions on the possible actions decided by two or more players, with a binding contract that forces the implementation of the agreement. Usually, to be binding an agreement requires an outside authority that can monitor the agreement at no cost and impose on violators sanctions so severe that cheating is prevented.

2.4 Games in Normal Form

2.4.1 Playing games through strategies

Let $M = \{1, \ldots, m\}$ be the set of players. A *pure strategy* γ_j for Player j is a mapping which transforms the information available to Player j at a decision node where he is making a move into his set of admissible actions. We call *strategy vector* the m-tuple $\gamma = (\gamma)_{j=1,\ldots m}$. Once a strategy is selected by each player, the strategy vector γ is defined and the game is played as it were controlled by an automaton[4].

An outcome (expressed in terms of expected utility to each player if the game includes chance nodes) is associated with a strategy vector γ. We denote by Γ_j the set

[4]This idea of playing games through the use of automata will be discussed in more details when we present the *folk theorem* for repeated games in Part II

of strategies for Player j. Then the game can be represented by the m mappings

$$V_j : \Gamma_1 \times \cdots \Gamma_j \times \cdots \Gamma_m \rightarrow \mathbf{R}, \quad j \in M$$

that associate a unique (expected utility) outcome $V_j(\gamma)$ for each player $j \in M$ with a given strategy vector in $\gamma \in \Gamma_1 \times \cdots \Gamma_j \times \cdots \Gamma_m$. One then says that the game is defined in its *normal form* .

2.4.2 From the extensive form to the strategic or normal form

We consider a simple two-player game, called "matching pennies". The rules of the game are as follows:

> The game is played over two stages. At first stage each player chooses head (H) or tail (T) without knowing the other player's choice. Then they reveal their choices to one another. If the coins do not match, Player 1 wins \$5 and Payer 2 wins -\$5. If the coins match, Player 2 wins \$5 and Payer 1 wins -\$5. At the second stage, the player who lost at stage 1 has the choice of either stopping the game or playing another penny matching with the same type of payoffs as in the first stage (Q, H, T).

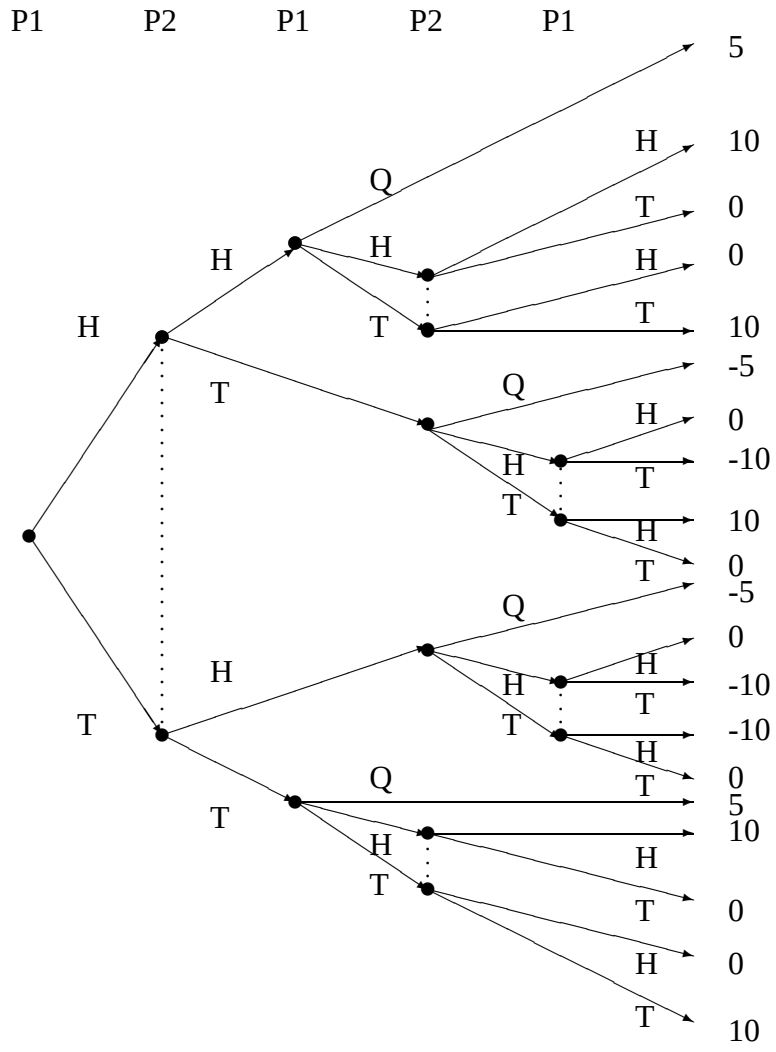
The extensive form tree

This game is represented in its extensive form in Figure 2.3. The terminal payoffs represent what Player 1 wins; Player 2 receives the opposite values. We have represented the information structure in a slightly different way here. A dotted line connects the different nodes forming an information set for a player. The player who has the move is indicated on top of the graph.

Listing all strategies

In Table 2.1 we have identified the 12 different strategies that can be used by each of the two players in the game of Matching pennies. Each player moves twice. In the first move the players have no information; in the second move they know what have been the choices made at first stage. We can easily identify the whole set of possible strategies.

Figure 2.3: The extensive form tree of the matching pennies game



Payoff matrix

In Table 2.2 we have represented the payoffs that are obtained by Player 1 when both players choose one of the 12 possible strategies.

	Strategies of Player 1			Strategies of Player 2		
	1st move	scnd move if player 2 has played		1st move move	scnd move if player 1 has played	
		H	T		H	T
1	H	Q	H	H	H	Q
2	H	Q	T	H	T	Q
3	H	H	H	H	H	H
4	H	H	T	H	T	H
5	H	T	H	H	H	T
6	H	T	T	H	T	T
7	T	H	Q	T	Q	H
8	T	T	Q	T	Q	T
9	T	H	H	T	H	H
10	T	T	H	T	H	T
11	T	H	T	T	T	H
12	T	T	T	T	T	T

Table 2.1: List of strategies

2.4.3 Mixed and Behavior Strategies

Mixing strategies

Since a player evaluates outcomes according to his VNM-utility functions he can envision to "mix" strategies by selecting one of them randomly, according to a lottery that he will define. This introduces one supplementary chance move in the game description.

For example, if Player j has p pure strategies $\gamma_{jk}, k = 1, \ldots, p$ he can select the strategy he will play through a lottery which gives a probability x_{jk} to the pure strategy γ_{jk}, $k = 1, \ldots, p$. Now the possible choices of action by Player j are elements of the set of all the probability distributions

$$\mathcal{X}_j = \{\mathbf{x}_j = (x_{jk})_{k=1,\ldots,p} | x_{jk} \geq 0, \quad \sum_{k=1}^{p} x_{jk} = 1.$$

We note that the set $\mathcal{X}_j$ is compact and convex in $\mathbf{R}^p$.

	1	2	3	4	5	6	7	8	9	10	11	12
1	-5	-5	-5	-5	-5	-5	-5	-5	0	0	10	10
2	-5	-5	-5	-5	-5	-5	5	5	10	10	0	0
3	0	-10	0	-10	-10	0	5	5	0	0	10	10
4	-10	0	-10	0	-10	0	5	5	10	10	0	0
5	0	-10	0	-10	0	-10	5	5	0	0	10	10
6	0	-10	0	-10	0	-10	5	5	10	10	0	0
7	5	5	0	0	10	10	-5	-5	-5	-5	-5	-5
8	5	5	10	10	0	0	-5	-5	-5	5	5	5
9	5	5	0	0	10	10	-10	0	-10	0	-10	0
10	5	5	10	10	0	0	-10	0	-10	0	-10	0
11	5	5	0	0	10	10	0	-10	0	-10	0	-10
12	5	5	10	10	0	0	0	-10	0	-10	0	-10

Table 2.2: Payoff matrix

Behavior strategies

A *behavior strategy* is defined as a mapping which associates with the information available to Player j at a decision node where he is making a move a probability distribution over his set of actions.

The difference between *mixed* and *behavior* strategies is subtle. In a mixed strategy, the player considers the set of possible strategies and picks one, at random, according to a carefully designed lottery. In a behavior strategy the player designs a strategy that consists in deciding at each decision node, according to a carefully designed lottery, this design being contingent to the information available at this node. In summary we can say that a behavior strategy is a strategy that includes randomness at each decision node. A famous theorem [Kuhn, 1953], that we give without proof, establishes that these two ways of introding randomness in the choice of actions are equivalent in a large class of games.

Theorem 2.4.1 *In an extensive game of perfect recall all mixed strategies can be represented as behavior strategies.*

Chapter 3

Solution concepts for noncooperative games

3.1 introduction

In this chapter we consider games described in their normal form and we propose different solution concept under the assumption that the players are non cooperating. In noncooperative games the players do not communicate between each other, they don't enter into negotiation for achieving a common course of action. They know that they are playing a game. They know how the actions, their own and the actions of the other players, will determine the payoffs of every player. However they will not cooperate.

To speak of a solution concept for a game one needs, first of all, to describe the game in its normal form. The solution of an m-player game will thus be a set of strategy vectors γ that have attractive properties expressed in terms of the payoffs received by the players.

Recall that an m-person game in *normal form* is defined by the following data

$$\{M, (\Gamma_i), (V_j) \text{ for } j \in M\},$$

where M is the set of players, $M = \{1, 2, ..., m\}$, and for each player $j \in M$, Γ_j is the set of strategies (also called the *strategy space*) and V_j, $j \in M$, is the payoff function that assigns a real number $V_j(\gamma)$ with a *strategy vector* $\gamma \in \Gamma_1 \times \Gamma_2 \times \ldots \times \Gamma_m$.

In this chapter we shall study different classes of games in normal form. The first category consists in the so-called *matrix games* describing a situation where two players are in a complete antagonistic situation since what a player gains the other

player looses, and where each player has a finite choice of strategies. Matrix games are also called *two player zero-sum finite games*. The second category will consist of two player games, again with a finite strategy set for each player, but where the payoffs are not zero-sum. These are the *nonzero-sum matrix games* or *bimatrix games*. The third category, will be the so-called *concave games* that encompass the previous classes of matrix and bimatrix games and for which we will be able to prove nice existence, uniqueness and stability results for a noncooperative game solution concept called *equilibrium*.

3.2 Matrix Games

Definition 3.2.1 *A game is zero-sum if the sum of the players' payoffs is always zero. Otherwise the game is nonzero-sum. A two-player zero-sum game is also called a duel.*

Definition 3.2.2 *A two-player zero-sum game in which each player has only a finite number of actions to choose from is called a matrix game.*

Let us explore how matrix games can be "solved". We number the players 1 and 2 respectively. Conventionally, Player 1 is the maximizer and has m (pure) strategies, say $i = 1, 2, ..., m$, and Player 2 is the minimizer and has n strategies to choose from, say $j = 1, 2, ..., n$. If Player 1 chooses strategy i while Player 2 picks strategy j, then Player 2 pays Player 1 the amount a_{ij}[1]. The set of all possible payoffs that Player 1 can obtain is represented in the form of the $m \times n$ matrix A with entries a_{ij} for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$. Now, the element in the $i-$th row and $j-$th column of the matrix A corresponds to the amount that Player 2 will pay Player 1 if the latter chooses strategy i and the former chooses strategy j. Thus one can say that in the game under consideration, Player 1 (the maximizer) selects rows of A while Player 2 (the minimizer) selects columns of that matrix, and as the result of the play, Player 2 pays Player 1 the amount of money specified by the element of the matrix in the selected row and column.

Example 3.2.1 *Consider a game defined by the following matrix:*

$$\begin{bmatrix} 3 & 1 & 8 \\ 4 & 10 & 0 \end{bmatrix}$$

[1]Negative payments are allowed. We could have said also that Player 1 receives the amount a_{ij} and Player 2 receives the amount $-a_{ij}$.

The question now is what can be considered as players' best strategies.

One possibility is to consider the players' *security levels*. It is easy to see that if Player 1 chooses the first row, then, whatever Player 2 does, Player 1 will get a payoff equal to at least 1 (*util*[2]). By choosing the second row, on the other hand, Player 1 risks getting 0. Similarly, by choosing the first column Player 2 ensures that he will not have to pay more than 4, while the choice of the second or third column may cost him 10 or 8, respectively. Thus we say that Player 1's *security level* is 1 which is ensured by the choice of the first row, while Player 2's security level is 4 and it is ensured by the choice of the first column. Notice that

$$1 = \max_i \min_j a_{ij}$$

and

$$4 = \min_j \max_i a_{ij}$$

which is the reason why the strategy which ensures that Player 1 will get at least the payoff equal to his security level is called his *maximin strategy*. Symmetrically, the strategy which ensures that Player 2 will not have to pay more than his security level is called his *minimax strategy*.

Lemma 3.2.1 *The following inequality holds*

$$\max_i \min_j a_{ij} \leq \min_j \max_i a_{ij}. \tag{3.1}$$

Proof: The proof of this result is based on the remark that, since both security levels are achievable, they necessarily satisfy the inequality (3.1). We give below a more detailed proof.

We note the obvious set of inequalities

$$\min_j a_{kj} \leq a_{kl} \leq \max_i a_{il} \tag{3.2}$$

which holds for all possible k and l. More formally, let (i^*, j^*) and (i_*, j_*) be defined by

$$a_{i^*j^*} = \max_i \min_j a_{ij} \tag{3.3}$$

and

$$a_{i_*j_*} = \min_j \max_i a_{ij} \tag{3.4}$$

[2]A util is the utility unit.

respectively. Now consider the payoff $a_{i^*j_*}$. Then, by (3.2) with $k = i^*$ and $l = j_*$ we get

$$\max_i(\min_j a_{ij}) \leq a_{i^*j_*} \leq \min_j(\max_i a_{ij}).$$

QED.

An important observation is that if Player 1 has to move first and Player 2 acts having seen the move made by Player 1, then the maximin strategy is Player 1's best choice which leads to the payoff equal to 1. If the situation is reversed and it is Player 2 who moves first, then his best choice will be the minimax strategy and he will have to pay 4. Now the question is what happens if the players move simultaneously. The careful study of the example shows that when the players move simultaneously the minimax and maximin strategies are not satisfactory "solutions" to this game. Notice that the players may try to improve their payoffs by anticipating each other's strategy. In the result of that we will see a process which in this case will not converge to any solution.

Consider now another example.

Example 3.2.2 *Let the matrix game A be given as follows*

$$\begin{bmatrix} 10 & -15^* & 20 \\ 20 & -30 & 40 \\ 30 & -45 & 60 \end{bmatrix}.$$

Can we find satisfactory strategy pairs?

It is easy to see that

$$\max_i \min_j a_{ij} = max\{-15, -30, -45\} = -15$$

and

$$\min_j \max_i a_{ij} = min\{30, -15, 60\} = -15$$

and that the pair of maximin and minimax strategies is given by

$$(i, j) = (1, 2).$$

That means that Player 1 should choose the first row while Player 2 should select the second column, which will lead to the payoff equal to -15.$\diamond$

In the above example, we can see that the players' maximin and minimax strategies "solve" the game in the sense that the players will be best off if they use these strategies.

3.2.1 Saddle-Points

Let us explore in more depth this class of strategies that solve the zero-sum matrix game.

Definition 3.2.3 *If in a matrix game $A = [a_{ij}]_{i=1,\ldots,m;j=1,\ldots,n;}$ there exists a pair (i^*, j^*) such that, for all $i1, \ldots, m$ and $j1, \ldots, n$*

$$a_{ij^*} \leq a_{i^*j^*} \leq a_{i^*j} \tag{3.5}$$

we say that the pair (i^*, j^*) is a saddle point in pure strategies for the matrix game.

As an immediate consequence, we see that, at a saddle point of a zero-sum game, the security levels of the two players are equal, *i.e.* ,

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{i^*j^*}.$$

What is less obvious is the fact that, if the security levels are equal then there exists a saddle point.

Lemma 3.2.2 *If, in a matrix game, the following holds*

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = v$$

then the game admits a saddle point in pure strategies.

Proof: Let i^* and j^* be a strategy pair that yields the security level payoffs v (resp. $-v$) for Player 1 (resp. Player 2). We thus have for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$

$$a_{i^*j} \geq \min_j a_{i^*j} = \max_i \min_j a_{ij} \tag{3.6}$$

$$a_{ij^*} \leq \max_i a_{ij^*} = \min_j \max_i a_{ij}. \tag{3.7}$$

Since

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{i^*j^*} = v$$

by (3.6)-(3.7) we obtain

$$a_{ij^*} \leq a_{i^*j^*} \leq a_{i^*j}$$

which is the saddle point condition. QED.

Saddle point strategies provide a solution to the game problem even if the players move simultaneously. Indeed, in Example 3.2.2, if Player 1 expects Player 2 to choose

the second column, then the first row will be his optimal choice. On the other hand, if Player 2 expects Player 1 to choose the first row, then it will be optimal for him to choose the second column. In other words, neither player can gain anything by unilaterally deviating from his saddle point strategy. Each strategy constitutes the best reply the player can have to the strategy choice of his opponent. This observation leads to the following definition.

Remark 3.2.1 *Using strategies i^* and j^*, Players 1 and 2 cannot improve their payoff by unilaterally deviating from $(i)^*$ or $((j)^*)$ respectively. We call such strategies an equilibrium.*

Saddle point strategies, as shown in Example 3.2.2, lead to both an *equilibrium* and a pair of *guaranteed payoffs*. Therefore such a strategy pair, if it exists, provides a solution to a matrix game, which is "good" in that rational players are likely to adopt this strategy pair.

3.2.2 Mixed strategies

We have already indicated in chapter 2 that a player could "mix" his strategies by resorting to a lottery to decide what to play. A reason to introduce mixed strategies in a matrix game is to enlarge the set of possible choices. We have noticed that, like in Example 3.2.1, many matrix games do not possess saddle points in the class of pure strategies. However Von Neumann proved that saddle point strategy pairs exist in the class of *mixed strategies*. Consider the matrix game defined by an $m \times n$ matrix A. (As before, Player 1 has m strategies, Player 2 has n strategies). A mixed strategy for Player 1 is an m-tuple

$$x = (x_1, x_2, ..., x_m)$$

where x_i are nonnegative for $i = 1, 2, ..., m$, and $x_1 + x_2 + ... + x_m = 1$. Similarly, a mixed strategy for Player 2 is an n-tuple

$$y = (y_1, y_2, ..., y_n)$$

where y_j are nonnegative for $j = 1, 2, ..., n$, and $y_1 + y_2 + ... + y_m = 1$.

Note that a pure strategy can be considered as a particular mixed strategy with one coordinate equal to one and all others equal to zero. The set of possible mixed strategies of Player 1 constitutes a simplex in the space $\mathbf{R}^m$. This is illustrated in Figure 3.1 for $m = 3$. Similarly the set of mixed strategies of Player 2 is a simplex in $\mathbf{R}^m$. A simplex is, by construction, the smallest closed convex set that contains $n + 1$ points in $\mathbf{R}^n$.

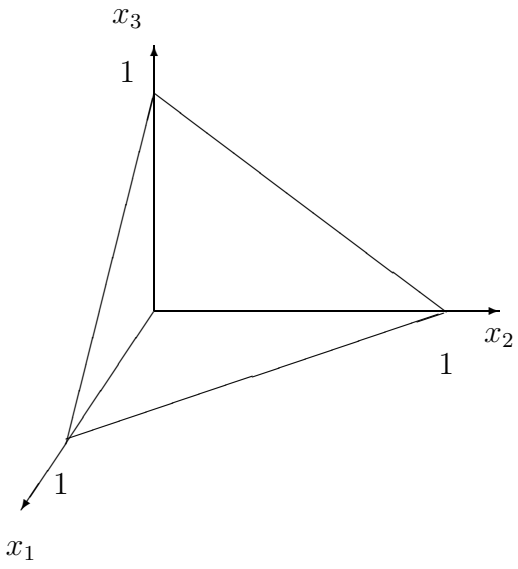


Figure 3.1: The simplex of mixed strategies

The interpretation of a mixed strategy, say x, is that Player 1, chooses his pure strategy i with probability $x_i,\ i = 1, 2, ..., m$. Since the two lotteries defining the random draws are independent events, the joint probability that the strategy pair (i, j) be selected is given by $x_i\, y_j$. Therefore, with each pair of mixed strategies (x, y) we can associate an expected payoff given by the quadratic expression (in $x.\ y$)[3]

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j a_{ij} = x^T A y.$$

The first important result of game theory proved in [Von Neumann 1928] showed the following

Theorem 3.2.1 *Any matrix game has a saddle point in the class of mixed strategies, i.e. , there exist probability vectors x and y such that*

$$\max_x \min_y x^T A y = \min_y \max_x x^T A y = (x^*)^T A y^* = v^*$$

where v^* is called the value of the game.

[3]The superscript ^{T} denotes the transposition operator on a matrix.

We shall not repeat the complex proof given by von Neumann. Instead we shall relate the search for saddle points with the solution of linear programs. A well known duality property will give us the saddle point existence result.

3.2.3 Algorithms for the Computation of Saddle-Points

Matrix games can be solved as linear programs. It is easy to show that the following two relations hold:

$$v^* = \max_x \min_y x^T A y = \max_x \min_j \sum_{i=1}^{m} x_i a_{ij} \tag{3.8}$$

and

$$z^* = \min_y \max_x x^T A y = \min_y \max_i \sum_{j=1}^{n} y_j a_{ij} \tag{3.9}$$

These two relations imply that the value of the matrix game can be obtained by solving any of the following two linear programs:

1. Primal problem

$$\begin{aligned} \max \quad & v \\ \text{subject to} \quad & \\ v \quad &\leq \quad \sum_{i=1}^{m} x_i a_{ij}, \quad j = 1, 2, ..., n \\ 1 \quad &= \quad \sum_{i=1}^{m} x_i \\ x_i \quad &\geq \quad 0, \quad i = 1, 2, ...m \end{aligned}$$

2. Dual problem

$$\begin{aligned} \min \quad & z \\ \text{subject to} \quad & \\ z \quad &\geq \quad \sum_{j=1}^{n} y_i a_{ij}, \quad i = 1, 2, ..., m \\ 1 \quad &= \quad \sum_{j=1}^{n} y_j \\ y_j \quad &\geq \quad 0, \quad j = 1, 2, ...n \end{aligned}$$

The following theorem relates the two programs together.

Theorem 3.2.2 *(Von Neumann [Von Neumann 1928]): Any finite two-person zero-sum matrix game A has a value*

proof: The value v^* of the zero-sum matrix game A is obtained as the common optimal value of the following pair of dual linear programming problems. The respective optimal programs define the saddle-point mixed strategies.

Primal		Dual	
max	v	min	z
subject to	$x^T A \geq v\mathbf{1}^T$	subject to	$A\,y \leq z\mathbf{1}$
	$x^T\mathbf{1} = 1$		$\mathbf{1}^T y = 1$
	$x \geq 0$		$y \geq 0$

where

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

denotes a vector of appropriate dimension with all components equal to 1. One needs to solve only one of the programs. The primal and dual solutions give a pair of saddle point strategies. •

Remark 3.2.2 *Simple $n \times n$ games can be solved more easily (see [Owen, 1982]). Suppose A is an $n \times n$ matrix game which does not have a saddle point in pure strategies. The players' unique saddle point mixed strategies and the game value are given by:*

$$x = \frac{\mathbf{1}^T A^D}{\mathbf{1}^T A^D \mathbf{1}} \tag{3.10}$$

$$y = \frac{A^D \mathbf{1}}{\mathbf{1}^T A^D \mathbf{1}} \tag{3.11}$$

$$v = \frac{det A}{\mathbf{1}^T A^D \mathbf{1}} \tag{3.12}$$

where A^D is the adjoint matrix of A, $det\ A$ the determinant of A, and $\mathbf{1}$ the vector of ones as before.

Let us illustrate the usefulness of the above formulae on the following example.

Example 3.2.3 *We want to solve the matrix game*

$$\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}.$$

The game, obviously, has no saddle point (in pure strategies). The adjoint A^D *is*

$$\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}$$

and $\mathbf{1}A^D = [3 \;\; 1]$, $A^D\mathbf{1}^T = [2 \;\; 2]$, $\mathbf{1}A^D\mathbf{1}^T = 4$, $det\ A = 2$. *Hence the best* mixed strategies *for the players are:*

$$x = [\frac{3}{4}, \;\frac{1}{4}], \quad y = [\frac{1}{2}, \;\frac{1}{2}]$$

and the value of the play is:

$$v = \frac{1}{2}$$

In other words in the long run Player 1 *is supposed to win* $.5$ *if he uses the first row* 75% *of times and the second* 25% *of times. The Player* 2*'s best strategy will be to use the first and the second column* 50% *of times which ensures him a loss of* $.5$ *(only; using other strategies he is supposed to lose more).*

3.3 Bimatrix Games

We shall now extend the theory to the case of a nonzero sum game. A *bimatrix game* conveniently represents a two-person nonzero sum game where each player has a finite set of possible pure strategies.

In a bimatrix game there are two players, say Player 1 and Player 2 who have m and n pure strategies to choose from respectively. Now, if the players select a pair of pure strategies, say (i, j), then Player 1 obtains the payoff a_{ij} and Player 2 obtains b_{ij}, where a_{ij} and b_{ij} are some given numbers. The payoffs for the two players corresponding to all possible combinations of pure strategies can be represented by two $m \times n$ payoff matrices A and B with entries a_{ij} and b_{ij} respectively (from here the name).

Notice that a (zero-sum) matrix game is a bimatrix game where $b_{ij} = -a_{ij}$. When $a_{ij} + b_{ij} = 0$, the game is a zero-sum matrix game. Otherwise, the game is nonzero-sum. (As a_{ij} and b_{ij} are the players' payoff this conclusion agrees with Definition 3.2.1.)

Example 3.3.1 *Consider the bimatrix game defined by the following matrices*

$$\begin{bmatrix} 52 & 44 & 44 \\ 42 & 46 & 39 \end{bmatrix}$$

and

$$\begin{bmatrix} 50 & 44 & 41 \\ 42 & 49 & 43 \end{bmatrix}.$$

It is often convenient to combine the data contained in two matrices and write it in the form of one matrix whose entries are ordered pairs (a_{ij}, b_{ij}). In this case one obtains

$$\begin{bmatrix} (52,50)^* & (44,44) & (44,41) \\ (42,42) & (46,49)^* & (39,43) \end{bmatrix}.$$

In the above bimatrix some cells have been indicated with asterisks $*$. They correspond to outcomes resulting from *equilibra* , a concept that we shall introduce now.

If Player 1 chooses strategy "row 1", the best reply by Player 2 is to choose strategy "column 1", and vice versa. Therefore we say that the outcome (52, 50) is associated with the equilibrium pair "row 1,column 1".

The situation is the same for the outcome (46, 49) that is associated with another equilibrium pair "row 2,column 2".

We already see on this simple example that a bimatrix game can have many equilibria. However there are other examples where no equilibrium can be found in pure strategies. So, as we have already done with (zero-sum) matrix games, let us expand the strategy sets to include mixed strategies.

3.3.1 Nash Equilibria

Assume that the players may use *mixed strategies* . For zero-sum matrix games we have shown the existence of a saddle point in mixed strategies which exhibits equilibrium properties. We formulate now the *Nash equilibrium* concept for the bimatrix games. The same concept will be defined later on for a more general m-player case.

Definition 3.3.1 *A pair of mixed strategies (x^*, y^*) is said to be a Nash equilibrium of the bimatrix game if*

1. $(x^*)^T A(y^*) \geq x^T A(y^*)$ *for every mixed strategy x, and*
2. $(x^*)^T B(y^*) \geq (x^*)^T By$ *for every mixed strategy y.*

In an equilibrium, no player can improve his payoff by deviating unilaterally from his equilibrium strategy.

Remark 3.3.1 *A Nash equilibrium extends to nonzero sum games the equilibrium property that was observed for the saddle point solution to a zero sum matrix game. The big difference with the saddle point concept is that, in a nonzero sum context, the equilibrium strategy for a player does not guarantee him that he will receive at least the equilibrium payoff. Indeed if his opponent does not play "well", i.e. does not use the equilibrium strategy, the outcome of a player can be anything; there is no guarantee.*

Another important step in the development of the theory of games has been the following theorem [Nash, 1951]

Theorem 3.3.1 *Every finite bimatrix game has at least one Nash equilibrium in* mixed strategies .

Proof: A more general existence proof including the case of bimatrix games will be given in the next section.•

3.3.2 Shortcommings of the Nash equilibrium concept

Multiple equilibria

As noticed in Example 3.3.1 a bimatrix game may have several equilibria in pure strategies. There may be additional equilibria in mixed strategies as well. The nonuniqueness of Nash equilibria for bimatrix games is a serious theoretical and practical problem. In Example 3.3.1 one equilibrium strictly *dominates* the other, ie., gives both players higher payoffs. Thus, it can be argued that even without any consultations the players will naturally pick the strategy pair $(i, j) = (1, 1)$.

It is easy to define examples where the situation is not so clear.

Example 3.3.2 *Consider the following bimatrix game*

$$\begin{bmatrix} (2,1)^* & (0,0) \\ (0,0) & (1,2)^* \end{bmatrix}$$

It is easy to see that this game has two equilibria (in pure strategies), none of which dominates the other. Moreover, Player 1 *will obviously prefer the solution* $(1, 1)$*, while Player* 2 *will rather have* $(2, 2)$*. It is difficult to decide how this game should be played if the players are to arrive at their decisions independently of one another.*

The prisoner's dilemma

There is a famous example of a bimatrix game, that is used in many contexts to argue that the Nash equilibrium solution is not always a "good" solution to a noncooperative game.

Example 3.3.3 *Suppose that two suspects are held on the suspicion of committing a serious crime. Each of them can be convicted only if the other provides evidence against him, otherwise he will be convicted as guilty of a lesser charge. By agreeing to give evidence against the other guy, a suspect can shorten his sentence by half. Of course, the prisoners are held in separate cells and cannot communicate with each other. The situation is as described in Table 3.1 with the entries giving the length of the prison sentence for each suspect, in every possible situation. In this case, the players are assumed to minimize rather than maximize the outcome of the play. The unique*

Suspect I:	Suspect II: refuses	agrees to testify
refuses	$(2, 2)$	$(10, 1)$
agrees to testify	$(1, 10)$	$(5, 5)^*$

Table 3.1: The Prisoner's Dilemma.

Nash equilibrium of this game is given by the pair of pure strategies $(agree - to - testify,\ agree - to - testify)$ *with the outcome that both suspects will spend five years in prison. This outcome is strictly dominated by the strategy pair* $(refuse - to - testify,\ refuse - to - testify)$*, which however is not an equilibrium and thus is not a realistic solution of the problem when the players cannot make binding agreements.*

This example shows that Nash equilibria could result in outcomes being very far from efficiency.

3.3.3 Algorithms for the Computation of Nash Equilibria in Bimatrix Games

Linear programming is closely associated with the characterization and computation of saddle points in matrix games. For bimatrix games one has to rely to algorithms solving either quadratic programming or complementarity problems. There are also a few algorithms (see [Aumann, 1989], [Owen, 1982]) which permit us to find an equilibrium of simple bimatrix games. We will show one for a 2×2 bimatrix game and then introduce the quadratic programming [Mangasarian and Stone, 1964] and complementarity problem [Lemke & Howson 1964] formulations.

Equilibrium computation in a 2×2 bimatrix game

For a simple 2×2 bimatrix game one can easily find a mixed strategy equilibrium as shown in the following example.

Example 3.3.4 *Consider the game with payoff matrix given below.*

$$\begin{bmatrix} (1,0) & (0,1) \\ (\frac{1}{2},\frac{1}{3}) & (1,0) \end{bmatrix}$$

Notice that this game has no pure strategy equilibrium. Let us find a mixed strategy equilibrium.

Assume Player 2 *chooses his equilibrium strategy* y *(i.e.* $100y\%$ *of times use first column,* $100(1-y)\%$ *times use second column) in such a way that Player* 1 *(in equilibrium) will get as much payoff using first row as using second row i.e.*

$$y + 0(1-y) = \frac{1}{2}y + 1(1-y).$$

This is true for $y^* = \frac{2}{3}$.

Symmetrically, assume Player 1 *is using a strategy* x *(i.e.* $100x\%$ *of times use first row,* $100(1-y)\%$ *times use second row) such that Player* 2 *will get as much payoff using first column as using second column i.e.*

$$0x + \frac{1}{3}(1-x) = 1x + 0(1-x).$$

This is true for $x^* = \frac{1}{4}$. *The players' payoffs will be, respectively,* $(\frac{2}{3})$ *and* $(\frac{1}{4})$.

Then the pair of mixed strategies

$$(x^*, 1-x^*),\ (y^*, 1-y^*)$$

is an equilibrium in mixed strategies.

Links between quadratic programming and Nash equilibria in bimatrix games

Mangasarian and Stone (1964) have proved the following result that links quadratic programming with the search of equilibria in bimatrix games. Consider a bimatrix

game (A, B). We associate with it the quadratic program

$$\begin{aligned}
\max \quad & [x^T Ay + x^T By - v^1 - v^2] && (3.13)\\
\text{s.t.} \quad & \\
Ay &\leq v^1 \mathbf{1}_m && (3.14)\\
B^T x &\leq v^2 \mathbf{1}_n && (3.15)\\
x, y &\geq 0 && (3.16)\\
x^T \mathbf{1}_m &= 1 && (3.17)\\
y^T \mathbf{1}_n &= 1 && (3.18)\\
v^1, v^2, &\in \mathbf{R}. && (3.19)
\end{aligned}$$

Lemma 3.3.1 *The following two assertions are equivalent*

(i) (x, y, v^1, v^2) *is a solution to the quadratic programming problem (3.13)- (3.19)*

(ii) (x, y) *is an equilibrium for the bimatrix game.*

Proof: From the constraints it follows that $x^T Ay \leq v^1$ and $x^T By \leq v^2$ for any feasible (x, y, v^1, v^2). Hence the maximum of the program is at most 0. Assume that (x, y) is an equilibrium for the bimatrix game. Then the quadruple

$$(x, y, v^1 = x^T Ay, v^2 = x^T By)$$

is feasible, i.e. satisfies (7.2)-(3.19), and gives to the objective function (7.1) a value 0. Hence the equilibrium defines a solution to the quadratic programming problem (7.1)-(3.19).

Conversely, let (x_*, y_*, v_*^1, v_*^2) be a solution to the quadratic programming problem (7.1)- (3.19). We know that an equilibrium exists for a bimatrix game (Nash theorem). We know that this equilibrium is a solution to the quadratic programming problem (7.1)-(3.19) with optimal value 0. Hence the optimal program (x_*, y_*, v_*^1, v_*^2) must also give a value 0 to the objective function and thus be such that

$$x_*^T Ay_* + x_*^T By_* \quad = \quad v_*^1 + v_*^2. \tag{3.20}$$

For any $x \geq 0$ and $y \geq 0$ such that $x^T \mathbf{1}_m = 1$ and $y^T \mathbf{1}_n = 1$ we have, by (3.17) and (3.18)

$$\begin{aligned}
x^T Ay_* &\leq v_*^1\\
x_*^T By &\leq v_*^2.
\end{aligned}$$

In particular we must have

$$\begin{aligned} x_*^T A y_* &\leq v_*^1 \\ x_*^T B y_* &\leq v_*^2 \end{aligned}$$

These two conditions with (3.20) imply

$$\begin{aligned} x_*^T A y_* &= v_*^1 \\ x_*^T B y_* &= v_*^2. \end{aligned}$$

Therefore we can conclude that For any $x \geq 0$ and $y \geq 0$ such that $x^T \mathbf{1}_m = 1$ and $y^T \mathbf{1}_n = 1$ we have, by (3.17) and (3.18)

$$\begin{aligned} x^T A y_* &\leq x_*^T A y_* \\ x_*^T B y &\leq x_*^T B y_* \end{aligned}$$

and hence, (x_*, y_*) is a Nash equilibrium for the bimatrix game. **QED**

A Complementarity Problem Formulation

We have seen that the search for equilibria could be done through solving a quadratic programming problem. Here we show that the solution of a bimatrix game can also be obtained as the solution of a *complementarity problem* .

There is no loss in generality if we assume that the payoff matrices are $m \times n$ and have only positive entries ($A > 0$ and $B > 0$). This is not restrictive, since VNM utilities are defined up to an increasing affine transformation. A strategy for Player 1 is defined as a vector $x \in \mathbf{R}^m$ that satisfies

$$x \geq 0 \tag{3.21}$$
$$x^T \mathbf{1}_m = 1 \tag{3.22}$$

and similarly for Player 2

$$y \geq 0 \tag{3.23}$$
$$y^T \mathbf{1}_n = 1. \tag{3.24}$$

It is easily shown that the pair (x^*, y^*) satisfying (3.21)-(3.24) is an equilibrium iff

$$\begin{aligned} ({x^*}^T A\, y^*)\mathbf{1}_m &\geq A\, y^* \qquad (A > 0) \\ ({x^*}^T B\, y^*)\mathbf{1}_n &\geq B^T x^* \qquad (B > 0) \end{aligned} \tag{3.25}$$

i.e. if the equilibrium condition is satisfied for pure strategy alternatives only.

Consider the following set of constraints with $v_1 \in \mathbf{R}$ and $v_2 \in \mathbf{R}$

$$\begin{array}{rcl} v_1\mathbf{1}_m & \geq & A\,y^* \\ v_2\mathbf{1}_n & \geq & B^T\,x^* \end{array} \quad \text{and} \quad \begin{array}{rcl} {x^*}^T(A\,y^* - v_1\mathbf{1}_m) & = & 0 \\ {y^*}^T(B^T\,x^* - v_2\mathbf{1}_n) & = & 0 \end{array}. \tag{3.26}$$

The relations on the right are called *complementarity constraints* . For mixed strategies (x^*, y^*) satisfying (3.21)-(3.24), they simplify to ${x^*}^T A\,y^* = v_1$, ${x^*}^T B\,y^* = v_2$. This shows that the above system (3.26) of constraints is equivalent to the system (3.25).

Define $s_1 = x/v_2$, $s_2 = y/v_1$ and introduce the slack variables u_1 and u_2, the system of constraints (3.21)-(3.24) and (3.26) can be rewritten

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_m \\ \mathbf{1}_n \end{pmatrix} - \begin{pmatrix} 0 & A \\ B^T & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \tag{3.27}$$

$$0 = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^T \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \tag{3.28}$$

$$0 \leq \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^T \tag{3.29}$$

$$0 \leq \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}. \tag{3.30}$$

Introducing four obvious new variables permits us to rewrite (3.27)- (3.30) in the generic formulation

$$u = q + Ms \tag{3.31}$$
$$0 = u^T s \tag{3.32}$$
$$u \geq 0 \tag{3.33}$$
$$s \geq 0, \tag{3.34}$$

of a so-called a *complementarity problem* .

A pivoting algorithm ([Lemke & Howson 1964], [Lemke 1965]) has been proposedto solve such problems. This algorithm applies also to *quadratic programming* , so this confirms that the solution of a bimatrix game is of the same level of difficulty as solving a quadratic programming problem.

Remark 3.3.2 *Once we obtain x and y, solution to (3.27)-(3.30) we shall have to reconstruct the strategies through the formulae*

$$x = s_1/(s_1^T\mathbf{1}_m) \tag{3.35}$$
$$y = s_2/(s_2^T\mathbf{1}_n). \tag{3.36}$$

3.4 Concave m-Person Games

The nonuniqueness of equilibria in bimatrix games and a fortiori in m-player matrix games poses a delicate problem. If there are many equilibria, in a situation where one assumes that the players cannot communicate or enter into preplay negotiations, how will a given player choose among the different strategies corresponding to the different equilibrium candidates? In single agent optimization theory we know that strict concavity of the (maximized) objective function and compactness and convexity of the constraint set lead to existence and uniqueness of the solution. The following question thus arises

> Can we generalize the mathematical programming approach to a situation where the optimization criterion is a Nash-Cournot equilibrium? can we then give sufficient conditions for existence and uniqueness of an equilibrium solution?

The answer has been given by Rosen in a seminal paper [Rosen, 1965] dealing with *concave m-person game* .

A *concave m-person game* is described in terms of individual strategies represented by vectors in compact subsets of Euclidian spaces ($\mathbf{R}^{m_j}$ for Player j) and by payoffs represented, for each player, by a continuous functions which is concave w.r.t. his own strategic variable. This is a generalization of the concept of mixed strategies, introduced in previous sections. Indeed, in a matrix or a bimatrix game the mixed strategies of a player are represented as elements of a *simplex* , *i.e.* a compact convex set, and the payoffs are *bilinear* or *multilinear* forms of the strategies, hence, for each player the payoff is concave w.r.t. his own strategic variable. This structure is thus generalized in two ways: (i) the strategies can be vectors constrained to be in a more general compact convex set and (ii) the payoffs are represented by more general continuous-concave functions.

Let us thus introduce the following game in strategic form

- Each player $j \in M = \{1, \ldots, m\}$ controls the action $u_j \in U_j$ a compact convex subset of $\mathbf{R}^{m_j}$, where m_j is a given integer, and gets a payoff $\psi_j(u_1, \ldots, u_j, \ldots, u_m)$, where $\psi_j : U_1 \times \ldots \times U_j, \ldots \times U_m \mapsto \mathbf{R}$ is continuous in each u_i and concave in u_j.

- A *coupled constraint* is defined as a proper subset $\mathcal{U}$ of $U_1 \times \ldots \times U_j \times \ldots \times U_m$. The constraint is that the joint action $\mathbf{u} = (u_1, \ldots, u_m)$ must be in $\mathcal{U}$.

Definition 3.4.1 *An equilibrium , under the coupled constraint $\mathcal{U}$ is defined as a decision m-tuple $(u_1^*, \ldots, u_j^*, \ldots, u_m^*) \in \mathcal{U}$ such that for each player $j \in M$*

$$\psi_j(u_1^*, \ldots, u_j^*, \ldots, u_m^*) \geq \psi_j(u_1^*, \ldots, u_j, \ldots, u_m^*) \tag{3.37}$$

$$\text{for all } u_j \in U_j \quad \text{s.t.} \quad (u_1^*, \ldots, u_j, \ldots, u_m^*) \in \mathcal{U}. \tag{3.38}$$

Remark 3.4.1 *The consideration of a coupled constraint is a new feature. Now each player's strategy space may depend on the strategy of the other players. This may look awkward in the context of nonocooperative games where thw player cannot enter into communication or cannot coordinate their actions. However the concept is mathematically well defined. We shall see later on that it fits very well some interesting aspects of environmental management. One can think for example of a global emission constraint that is imposed on a finite set of firms that are competing on the same market. This environmental example will be further developed in forthcoming chapters.*

3.4.1 Existence of Coupled Equilibria

Definition 3.4.2 *A coupled equilibrium is a vector $\mathbf{u}^*$ such that*

$$\psi_j(\mathbf{u}^*) = \max_{u_j} \{\psi_j(u_1^*, \ldots, u_j, \ldots, u_m^*) | (u_1^*, \ldots, u_j, \ldots, u_m^*) \in \mathcal{U}\}. \tag{3.39}$$

At such a point no player can improve his payoff by a unilateral change in his strategy which keeps the combined vector in $\mathcal{U}$.

Let us first show that an equilibrium is actually defined through a *fixed point* condition. For that purpose we introduce a so-called *global reaction function* $\theta : \mathcal{U} \times \mathcal{U} \mapsto \mathbf{R}$ defined by

$$\theta(\mathbf{u}, \mathbf{v}, \mathbf{r}) = \sum_{j=1}^{m} r_j \psi_j(u_1, \ldots, v_j, \ldots, u_m), \tag{3.40}$$

where $r_j > 0$, $j = 1, \ldots, m$ are given weights. The precise role of this weighting scheme will be explained later. For the moment we could take as well $r_j \equiv 1$. Notice that, even if $\mathbf{u}$ and $\mathbf{v}$ are in $\mathcal{U}$, the combined vectors $(u_1, \ldots, v_j, \ldots, u_m)$ are element of a larger set in $U_1 \times \ldots \times U_m$. This function is continuous in $\mathbf{u}$ and concave in $\mathbf{v}$ for every fixed $\mathbf{u}$. We call it a *reaction function* since the vector v can be interpreted as composed of the reactions of the different players to the given vector u. This function is helpful as shown in the following result.

Lemma 3.4.1 *Let $\mathbf{u}^* \in \mathcal{U}$ be such that*

$$\theta(\mathbf{u}^*, \mathbf{u}^*, \mathbf{r}) = \max_{\mathbf{u} \in \mathcal{U}} \theta(\mathbf{u}^*, \mathbf{u}, \mathbf{r}). \tag{3.41}$$

Then $\mathbf{u}^*$ is a coupled equilibrium.

Proof: Assume $\mathbf{u}^*$ satisfies (3.41) but is not a coupled equilibrium, *i.e.* does not satisfy (3.39). Then, for one player, say ℓ, there would exist a vector

$$\bar{\mathbf{u}} = (u_1^*, \ldots, u_\ell, \ldots, u_m^*) \in \mathcal{U}$$

such that

$$\psi_\ell(u_1^*, \ldots, u_\ell, \ldots, u_m^*) > \psi_\ell(\mathbf{u}^*).$$

Then we shall also have $\theta(\mathbf{u}^*, \bar{\mathbf{u}}) > \theta(\mathbf{u}^*, \mathbf{u}^*)$ which is a contradiction to (3.41). **QED**

This result has two important consequences.

1. It shows that proving the existence of an equilibrium reduces to proving that a *fixed point* exists for an appropriately defined *reaction mapping* ($\mathbf{u}^*$ is the best reply to $\mathbf{u}^*$ in (3.41));
2. it associates with an equilibrium an *implicit maximization problem* , defined in (3.41). We say that this problem is *implicit* since it is defined in terms of its solution $\mathbf{u}^*$.

To make more precise the fixed point argument we introduce a *coupled reaction mapping* .

Definition 3.4.3 *The point to set mapping*

$$\Gamma(\mathbf{u}, \mathbf{r}) = \{\mathbf{v} | \theta(\mathbf{u}, \mathbf{v}, \mathbf{r}) = \max_{\mathbf{w} \in \mathcal{U}} \theta(\mathbf{u}, \mathbf{w}, \mathbf{r})\}. \tag{3.42}$$

is called the coupled reaction mapping associated with the positive weighting $\mathbf{r}$. *A fixed point of* $\Gamma(\cdot, \mathbf{r})$ *is a vector* $\mathbf{u}^*$ *such that* $\mathbf{u}^* \in \Gamma(\mathbf{u}^*, \mathbf{r})$.

By Lemma 3.4.1 a fixed point of $\Gamma(\cdot, \mathbf{r})$ is a coupled equilibrium.

Theorem 3.4.1 *For any positive weighting* $\mathbf{r}$ *there exists a fixed point of* $\Gamma(\cdot, \mathbf{r})$*, i.e. a point* $\mathbf{u}^*$ *s.t.* $\mathbf{u}^* \in \Gamma(\mathbf{u}^*, \mathbf{r})$. *Hence a coupled equilibrium exists.*

Proof: The proof is based on the Kakutani fixed-point theorem that is given in the Appendix of section 3.8. One is required to show that the point to set mapping is upper semicontinuous. **QED**

Remark 3.4.2 *This existence theorem is very close, in spirit, to the theorem of Nash. It uses a fixed-point result which is "topological" and not "constructive", i.e. it does not provide a computational method. However, the definition of a* normalised equilibrium *introduced by Rosen establishes a link between mathematical programming and concave games with coupled constraints.*

3.4.2 Normalized Equilibria

Kuhn-Tucker multipliers

Suppose that the coupled constraint (3.39) can be defined by a set of inequalities

$$h_k(\mathbf{u}) \geq 0, \quad k = 1, \ldots, p \tag{3.43}$$

where $h_k : U_1 \times \ldots \times U_m \to \mathbf{R}$, $k = 1, \ldots, p$, are given concave functions. Let us further assume that the payoff functions $\psi_j(\cdot)$ as well as the constraint functions $h_k(\cdot)$ are continuously differentiable and satisfy the constraint qualification conditions so that Kuhn-Tucker multipliers exist for each of the implicit single agent optimization problems defined below.

Assume all players other than Player j use their strategy u_i^*, $i \in M$, $i \neq j$. Then the equilibrium conditions (3.37)-(3.39) define a single agent optimization problem with concave objective function and convex compact admissible set. Under the assumed constraint qualification assumption there exists a vector of Kuhn-Tucker multipliers $\lambda_j = (\lambda_{jk})_{k=1,\ldots,p}$ such that the Lagrangean

$$L_j([\mathbf{u}^{*-j}, u_j], \lambda_j) = \psi_j([\mathbf{u}^{*-j}, u_j]) + \sum_{k=1\ldots p} \lambda_{jk} h_k([\mathbf{u}^{*-j}, u_j]) \tag{3.44}$$

verifies, at the optimum

$$\begin{aligned} 0 &= \frac{\partial}{\partial u_j} L_j([\mathbf{u}^{*-j}, u_j^*], \lambda_j) && (3.45)\\ 0 &\leq \lambda_j && (3.46)\\ 0 &= \lambda_{jk} h_k([\mathbf{u}^{*-j}, u_j^*]) \quad k = 1, \ldots, p. && (3.47) \end{aligned}$$

Definition 3.4.4 *We say that the equilibrium is normalized if the different multipliers λ_j for $j \in M$ are colinear with a common vector λ_0, namely*

$$\lambda_j = \frac{1}{r_j} \lambda_0 \tag{3.48}$$

where $r_j > 0$ is a given weighting of the players.

Actually, this common multiplier λ_0 is associated with the implicit mathematical programming problem

$$\max_{\mathbf{u} \in \mathcal{U}} \theta(\mathbf{u}^*, \mathbf{u}, \mathbf{r}) \tag{3.49}$$

to which we associate the Lagrangean

$$L_0(\mathbf{u}, \lambda_0) = \sum_{j \in M} r_j \psi_j([\mathbf{u}^{*-j}, u_j]) + \sum_{k=1\ldots p} \lambda_{0k} h_k(\mathbf{u}). \tag{3.50}$$

and the first order necessary conditions

$$\begin{aligned} 0 &= \frac{\partial}{\partial u_j}\{r_j\psi_j(\mathbf{u}^*) + \sum_{k=1,\ldots,p} \lambda_{0k} h_k(\mathbf{u}^*)\}, \quad j \in M && (3.51)\\ 0 &\leq \lambda_0 && (3.52)\\ 0 &= \lambda_{0k} h_k(\mathbf{u}^*) \quad k = 1,\ldots,p. && (3.53) \end{aligned}$$

An economic interpretation

The multiplier, in a mathematical programming framework, can be interpreted as a marginal cost associated with the right-hand side of the constraint. More precisely it indicates the sensitivity of the optimum solution to marginal changes in this right-hand-side. The multiplier permits also a *price decentralization* in the sense that, through an ad-hoc pricing mechanism the optimizing agent is induced to satisfy the constraints. In a normalized equilibrium, the shadow cost interpretation is not so apparent; however, the *price decomposition principle* is still valid. Once the common multiplier has been defined, with the associated weighting $r_j > 0$, $j = 1,\ldots,m$ the coupled constraint will be satisfied by equilibrium seeking players, when they use as payoffs the Lagrangeans

$$L_j([\mathbf{u}^{*-j}, u_j], \lambda_j) = \psi_j([\mathbf{u}^{*-j}, u_j]) + \frac{1}{r_j}\sum_{k=1..p} \lambda_{0k} h_k([\mathbf{u}^{*-j}, u_j]), \quad j = 1,\ldots,m. \qquad (3.54)$$

The common multiplier permits then an "implicit pricing" of the common constraint so that it remains compatible with the equilibrium structure. Indeed, this result to be useful necessitates uniqueness of the normalized equilibrium associated with a given weighting $r_j > 0$, $j = 1,\ldots,m$. In a mathematical programming framework, uniqueness of an optimum results from strict concavity of the objective function to be maximized. In a game structure, uniqueness of the equilibrium will result from a more stringent strict concavity requirement, called by Rosen *strict diagonal concavity* .

3.4.3 Uniqueness of Equilibrium

Let us consider the so-called *pseudo-gradient* defined as the vector

$$g(\mathbf{u}, \mathbf{r}) = \begin{pmatrix} r_1 \frac{\partial}{\partial u_1}\psi_1(\mathbf{u}) \\ r_2 \frac{\partial}{\partial u_2}\psi_2(\mathbf{u}) \\ \vdots \\ r_m \frac{\partial}{\partial u_m}\psi_m(\mathbf{u}) \end{pmatrix} \qquad (3.55)$$

We notice that this expression is composed of the partial gradients of the different payoffs with respect to the decision variables of the corresponding player. We also consider the function

$$\sigma(\mathbf{u}, \mathbf{r}) = \sum_{j=1}^{m} r_j \psi_j(\mathbf{u}) \tag{3.56}$$

Definition 3.4.5 *The function $\sigma(\mathbf{u}, \mathbf{r})$ is* **diagonally strictly concave** *on $\mathcal{U}$ if, for every $\mathbf{u}^1$ et $\mathbf{u}^2$ in $\mathcal{U}$, the following holds*

$$(\mathbf{u}^2 - \mathbf{u}^1)^T g(\mathbf{u}^1, \mathbf{r}) + (\mathbf{u}^1 - \mathbf{u}^2)^T g(\mathbf{u}^2, \mathbf{r}) > 0. \tag{3.57}$$

A sufficient condition that $\sigma(\mathbf{u}, \mathbf{r})$ be *diagonally strictly concave* is that the symmetric matrix $[G(\mathbf{u}, \mathbf{r}) + G(\mathbf{u}, \mathbf{r})^T]$ be negative definite for any $\mathbf{u}^1$ in $\mathcal{U}$, where $G(\mathbf{u}, \mathbf{r})$ is the Jacobian of $g(\mathbf{u}^0, \mathbf{r})$ with respect to $\mathbf{u}$.

Theorem 3.4.2 *If $\sigma(\mathbf{u}, \mathbf{r})$ is* **diagonally strictly concave** *on the convex set $\mathcal{U}$, with the assumptions insuring existence of K.T. multipliers, then for every $\mathbf{r} > 0$ there exists a unique normalized equilibrium*

Proof We sketch below the proof given by Rosen [Rosen, 1965]. Assume that for some $r > 0$ we have two equilibria $\mathbf{u}^1$ and $\mathbf{u}^2$. Then we must have

$$h(\mathbf{u}^1) \geq 0 \tag{3.58}$$
$$h(\mathbf{u}^2) \geq 0 \tag{3.59}$$

and there exist multipliers $\lambda^1 \geq 0$, $\lambda^2 \geq 0$, such that

$$\lambda^{1^T} h(\mathbf{u}^1) = 0 \tag{3.60}$$
$$\lambda^{2^T} h(\mathbf{u}^2) = 0 \tag{3.61}$$

and for which the following holds true for each player $j \in M$

$$r_j \nabla_{u_i} \psi_j(\mathbf{u}^1) + \lambda^{1^T} \nabla_{u_i} h(\mathbf{u}^1) = 0 \tag{3.62}$$
$$r_j \nabla_{u_i} \psi_j(\mathbf{u}^2) + \lambda^{2^T} \nabla_{u_i} h(\mathbf{u}^2) = 0. \tag{3.63}$$

We multiply (3.62) by $(\mathbf{u}^2 - \mathbf{u}^1)^T$ and (3.63) by $(\mathbf{u}^1 - \mathbf{u}^2)^T$ and we sum together to obtain an expression $\beta + \gamma = 0$, where, due to the concavity of the h_k and the conditions (3.58)-(3.61)

$$\begin{aligned}
\gamma &= \sum_{j \in M} \sum_{k=1}^{p} \{\lambda_k^1 (\mathbf{u}^2 - \mathbf{u}^1)^T \nabla_{u_i} h_k(\mathbf{u}^1) + \lambda_k^2 (\mathbf{u}^1 - \mathbf{u}^2)^T \nabla_{u_i} h_k(\mathbf{u}^2)\} \\
&\geq \sum_{j \in M} \{\lambda^{1^T} [h(\mathbf{u}^2) - h(\mathbf{u}^1)] + \lambda^{2^T} [h(\mathbf{u}^1) - h(\mathbf{u}^2)]\} \\
&= \sum_{j \in M} \{\lambda^{1^T} h(\mathbf{u}^2) + \lambda^{2^T} h(\mathbf{u}^1)\} \geq 0,
\end{aligned} \tag{3.64}$$

and

$$\beta = \sum_{j \in M} r_j[(\mathbf{u}^2 - \mathbf{u}^1)^T \nabla_{u_i} \psi_j(\mathbf{u}^1) + (\mathbf{u}^1 - \mathbf{u}^2)^T \nabla_{u_i} \psi_j(\mathbf{u}^2)]. \quad (3.65)$$

Since $\sigma(\mathbf{u}, r)$ is diagonally strictly concave we have $\beta > 0$ which contradicts $\beta + \gamma = 0$. **QED**

3.4.4 A numerical technique

The diagonal strict concavity property that yielded the uniqueness result of theorem 3.4.2 also provides an interesting extension of the gradient method for the computation of the equilibrium. The basic idea is to "project", at each step, the pseudo gradient $g(\mathbf{u}, \mathbf{r})$ on the constraint set $\mathcal{U} = \{\mathbf{u} : h(\mathbf{u}) \geq 0\}$ (let's call $\bar{g}(\mathbf{u}^\ell, \mathbf{r})$ this projection) and to proceed through the usual steepest ascent step

$$\mathbf{u}^{\ell+1} = \mathbf{u}^\ell + \tau^\ell \bar{g}(\mathbf{u}^\ell, \mathbf{r}).$$

Rosen shows that, at each step ℓ the step size $\tau^\ell > 0$ can be chosen small enough for having a reduction of the norm of the projected gradient. This yields convergence of the procedure toward the unique equilibrium.

3.4.5 A variational inequality formulation

Theorem 3.4.3 *Under assumptions 3.5.1 the vector* $\mathbf{q}^* = (q_1^*, \ldots, q_m^*)$ *is a Nash-Cournot equilibrium if and only if it satisfies the following* variational inequality

$$(\mathbf{q} - \mathbf{q}^*)^T g(\mathbf{q}^*) \leq 0, \quad (3.66)$$

where $g(\mathbf{q}^*)$ *is the pseudo-gradient at* $\mathbf{q}^*$ *with weighting* $\mathbf{1}$.

Proof: Apply first order necessary and sufficient optimality conditions for each player and aggregate to obtain (3.66). QED

Remark 3.4.3 *The diagonal strict concavity assumption is then equivalent to the property of* strong monotonicity *of the operator* $g(\mathbf{q}^*)$ *in the parlance of variational inequality theory.*

3.5 Cournot equilibrium

The model proposed in 1838 by [Cournot, 1838] is still one of the most frequently used game model in economic theory. It represents the competition between different firms supplying a market for a divisible good.

3.5.1 The static Cournot model

Let us first of all recall the basic Cournot oligopoly model. We consider a single market on which m firms are competing. The market is characterized by its (inverse) demand law $p = D(Q)$ where p is the market clearing price and $Q = \sum_{j=1,\dots,m} q_j$ is the total supply of goods on the market. The firm j faces a cost of production $C_j(q_j)$, hence, letting $q = (q_1, \dots, q_m)$ represent the production decision vector of the m firms together, the profit of firm j is $\pi_j(q) = q_j\, D(Q) - C_j(q_j)$. The following assumptions are placed on the model

Assumption 3.5.1 *The market demand and the firms cost functions satisfy the following (i) The inverse demand function is finite valued, nonnegative and defined for all $Q \in [0, \infty)$. It is also twice differentiable, with $D'(Q) < 0$ wherever $D(Q) > 0$. In addition $D(0) > 0$. (ii) $C_j(q_j)$ is defined for all $q_j \in [0, \infty)$, nonnegative, convex, twice continuously differentiable, and $C'(q_j) > 0$. (iii) $Q\, D(Q)$ is bounded and strictly concave for all Q such that $D(Q) > 0$.*

If one assumes that each firm j chooses a supply level $q_j \in [0, \bar{q}_j]$, this model satisfies the definition of a concave game à la Rosen (see exercise 3.7). There exists an equilibrium. Let us consider the uniqueness issue in the duopoly case.

Consider the pseudo gradient (there is no need of weighting (r_1, r_2) since the constraints are not coupled)

$$g(q_1, q_2) = \begin{pmatrix} D(Q) + q_1\, D'(Q) - C_1'(q_1) \\ D(Q) + q_2\, D'(Q) - C_2'(q_2) \end{pmatrix}$$

and the jacobian matrix

$$G(q_1, q_2) = \begin{pmatrix} 2D'(Q) + q_1\, D''(Q) - C_1''(q_1) & D'(Q) + q_1\, D''(Q) \\ D'(Q) + q_2\, D''(Q) & 2D'(Q) + q_2\, D''(Q) - C_1''(q_2) \end{pmatrix}.$$

The negative definiteness of the symmetric matrix

$$\frac{1}{2}[G(q_1, q_2) + G(q_1, q_2)^T] =$$
$$\begin{pmatrix} 2D'(Q) + q_1\, D''(Q) - C_1''(q_1) & D'(Q) + \frac{1}{2}Q\, D''(Q) \\ D'(Q) + \frac{1}{2}Q\, D''(Q) & 2D'(Q) + q_2\, D''(Q) - C_1''(q_2) \end{pmatrix}$$

implies uniqueness of the equilibrium.

3.5.2 Formulation of a Cournot equilibrium as a nonlinear complementarity problem

We have seen, when we dealt with Nash equilibria in m-matrix games, that a solution could be found through the solution of a linear complementarity problem. A similar formulation holds for a Cournot equilibrium and it will in general lead to a *nonlinear complementarity problem* or NLCP, a class of problems that has recently been the object of considerable developments in the field of Mathematical Programming (see e.g. the book [Ferris & Pang, 1997a] and the survey paper [Ferris & Pang, 1997b]) We can rewrite the equilibrium condition as an NLCP in canonical form

$$q_j \geq 0 \tag{3.67}$$
$$f_j(q) \geq 0 \tag{3.68}$$
$$q_j f_j(q) = 0, \tag{3.69}$$

where

$$f_j(q) = -D(Q) - q_j\, D'(Q) + C'_j(q_j). \tag{3.70}$$

The relations (3.67-3.69) express that, at a Cournot equilibrium, either $q_j = 0$ and the gradient $g_j(q)$ is ≤ 0, or $q_j > 0$ and $g_j(q) = 0$. There are now optimization softwares that solve these types of problems. In [Rutherford, 1995] an extension of the GAMS (see [Brooke et al., 1992]) modeling software to provide an ability to solve this type of complementarity problem is described. More recently the modeling language AMPL (see [Fourer et al., 1993]) has also been adapted to handel a problem of this type and to submit it to an efficient NLCP solver like PATH (see [Ferris & Munson, 1994]).

Example 3.5.1 *This example comes from the AMPL website*

`http://www.ampl.com/ampl/TRYAMPL/`

Consider an oligopoly with $n = 10$ firms. The production cost function of firm j is given by

$$c_i(q_i) = c_i q_i + \frac{\beta_i}{1+\beta_i}(L_i q_i)^{(1+\beta_i)/\beta_i}. \tag{3.71}$$

The market demand function is defined by

$$D(Q) = (A/Q)^{1/\gamma}. \tag{3.72}$$

Therefore

$$D'(Q) = \frac{1}{\gamma}(A/Q)^{1/\gamma-1}(-\frac{A}{Q^2}) = -\frac{1}{\gamma}D(Q)/Q. \tag{3.73}$$

The Nash Cournot equilibrium will be the solution of the NLCP (3.67-3.69) where

$$\begin{aligned} f_j(q) &= -D(Q) - q_j\, D'(Q) + C_j'(q_j) \\ &= -(A/Q)^{1/\gamma} + \frac{1}{\gamma} q_i D(Q)/Q + c_i + (L_i q_i)^{1/\beta_i}. \end{aligned}$$

AMPL input

```
%%%%%%%FILE nash.mod%%%%%%%%%%

#==>nash.gms

# A non-cooperative game example: a Nash equilibrium is sought.

# References:

# F.H. Murphy, H.D. Sherali, and A.L. Soyster, "A Mathematical
#       Programming Approach for Determining Oligopolistic Market
#       Equilibrium", Mathematical Programming 24 (1986), pp. 92-
106.

# P.T. Harker, "Accelerating the convergence . . .", Mathematical
#       Programming 41 (1988), pp. 29-59.

set Rn := 1 .. 10;

param gamma := 1.2;

param c {i in Rn};
param beta {i in Rn};
param L {i in Rn} := 10;

var q {i in Rn} >= 0;           # production vector
var Q = sum {i in Rn} q[i];
var divQ = (5000/Q)**(1/gamma);

s.t. feas {i in Rn}:
        q[i] >= 0 complements
        0 <= c[i] + (L[i] * q[i])**(1/beta[i]) - divQ
             - q[i] * (-1/gamma) * divQ / Q;

%%%%%%%FILE nash.dat%%%%%%%%%%
data;

param c :=
```

```
1 5
2 3
3 8
4 5
5 1
6 3
7 7
8 4
9 6
10 3
;

param beta :=
1 1.2
2 1
3 .9
4 .6
5 1.5
6 1
7 .7
8 1.1
9 .95
10 .75
;

%%%%%%%FILE nash.run%%%%%%%%%%

model nash.mod;
data nash.dat;

set initpoint := 1 .. 4;
param initval {Rn,initpoint} >= 0;
data;param initval
: 1 2 3 4 :=
1 1       10 1.0 7
2 1       10 1.2 4
3 1       10 1.4 3
4 1       10 1.6 1
5 1       10 1.8 18
6 1       10 2.1 4
7 1       10 2.3 1
8 1       10 2.5 6
9 1       10 2.7 3
10 1   10 2.9 2
;

for {point in initpoint}
```

```
{
let{i in Rn} q[i] := initval[i,point];
solve;
include compchk
}
```

3.5.3 Computing the solution of a classical Cournot model

There are many approaches to find a solution to a static Cournot game. We have described above the one based on a solution of an NLCP. We list below a few other alternatives

(a) one discretizes the decision space and uses the complementarity algorithm proposed in [Lemke & Howson 1964] to compute a solution to the associated bimatrix game;

(b) one exploits the fact that the players are uniquely coupled through the condition that $Q = \sum_{j=1,\ldots,m} q_j$ and a sort of primal-dual search technique is used [Murphy,Sherali & Soyster, 1982];

(c) one formulates the Nash-Cournot equilibrium problem as a variational inequality. Assumption 3.5.1 implies *Strict Diagonal Concavity (SDC)* , in the sense of Rosen, i.e. strict monotonicity of the pseudo-gradient operator, hence uniqueness and convergence of various gradient like algorithms.

3.6 Correlated equilibria

Aumann, [Aumann, 1974], has proposed a mechanism that permits the players to enter into preplay arrangements so that their strategy choices could be correlated, instead of being independently chosen, while keeping an equilibrium property. He called this type of solution *correlated equilibrium*.

3.6.1 Example of a game with correlated equlibria

Example 3.6.1 *This example has been initially proposed by Aumann [Aumann, 1974]. Consider a simple bimatrix game defined as follows*

	c_1	c_2
r_1	$5,1$	$0,0$
r_2	$4,4$	$1,5$

This game has two pure strategy equilibria (r_1, c_1) *and* (r_2, c_2) *and a mixed strategy equilibrium where each player puts the same probality 0.5 on each possible pure strategy. The respective outcomes are shown in Table 3.2 Now, if the players agree to play*

(r_1, c_1)	:	5,1
(r_2, c_2)	:	1,5
$(0.5 - 0.5, 0.5 - 0.5)$	:	2.5,2.5

Table 3.2: Outcomes of the three equilibria

by jointly observing a "coin flip" and playing (r_1, c_1) *if the result is "head",* (r_2, c_2) *if it is "tail" then they expect the outcome* $3, 3$ *which is a convex combination of the two pure equilibrium outcomes.*

By deciding to have a coin flip deciding on the equilibrium to play, a new type of equilibrium has been introduced that yields an outcome located in the convex hull of the set of Nash equilibrium outcomes of the bimatrix game. In Figure 3.2 we have represented these different outcomes. The three full circles represent the three Nash equilibrium outcomes. The triangle defined by these three points is the convex hull of the Nash equilibrium outcomes. The empty circle represents the outcome obtained by agreeing to play according to the coin flip mechanism.

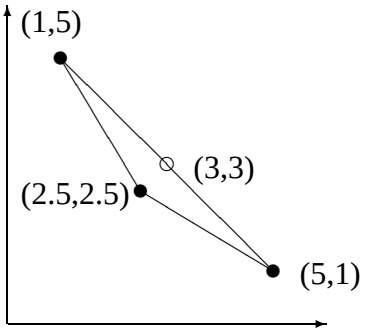


Figure 3.2: The convex hull of Nash equilibria

It is easy to see that this way to play the game defines an equilibrium for the extensive game shown in Figure 3.3. This is an expanded version of the initial game where, in a preliminary stage, "Nature" decides randomly the signal that will be observed by the players[4]*. The result of the "coin flip" is a "public information", in the sense that*

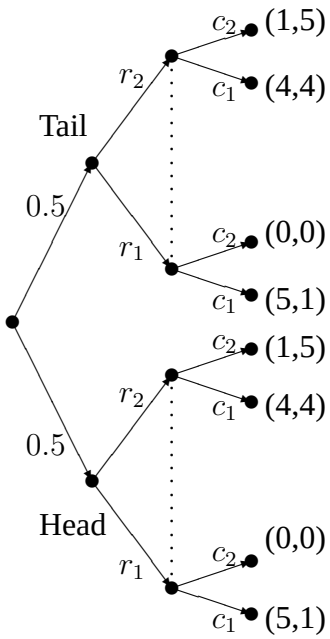


Figure 3.3: Extensive game formulation with "Nature" playing first

it is shared by all the players. One can easily check that the correlated equilibrium is a Nash equilibrium for the expanded game.

We now push the example one step further by assuming that the players agree to play according to the following mechanism: A random device selects one cell in the game matrix with the probabilities shown in Figure 3.74.

	c_1	c_2
r_1	$\frac{1}{3}$	0
r_2	$\frac{1}{3}$	$\frac{1}{3}$

(3.74)

When a cell is selected, then each player is told to play the corresponding pure strategy. The trick is that a player is told what to play but is not told what the recommendation to the other player is. The information received by each player is not "public" anymore. More precisely, the three possible signals are (r_1, c_1)*,* (r_2, c_1)*,* (r_2, c_2)*. When Player 1*

[4]Dotted lines represent information sets of Player 2.

receives the signal "play r_2" he knows that with a probability $\frac{1}{2}$ the other player has been told "play c_1" and with a probability $\frac{1}{2}$ the other player has been told "play c_2". When player 1 receives the signal "play r_1" he knows that with a probability 1 the other player has been told "play c_1". Consider now what Player 1 can do, if he assumes that the other player plays according to the recommendation. If Player 1 has been told "play r_2" and if he plays so he expects

$$\frac{1}{2} \times 4 + \frac{1}{2} \times 1 = 2.5;$$

if he plays r_1 instead he expects

$$\frac{1}{2} \times 5 + \frac{1}{2} \times 0 = 2.5$$

so he cannot improve his expected reward. If Player 1 has been told "play r_1" and if he plays so he expects 5, whereas if he plays r_2 he expects 4. So, for Player 1, obeying to the recommendation is the best reply to Player 2 behavior when he himself plays according to the suggestion of the signalling scheme. Now we can repeat the the verification for Player 2. If he has been told "play c_1" he expects

$$\frac{1}{2} \times 1 + \frac{1}{2} \times 4 = 2.5;$$

if he plays c_2 instead he expects

$$\frac{1}{2} \times 0 + \frac{1}{2} \times 5 = 2.5$$

so he cannot improve. If he has been told "play c_2" he expects 5 whereas if he plays c_1 instead he expects 4 so he is better off with the suggested play. So we have checked that an equilibrium property holds for this way of playing the game. All in all, each player expects

$$\frac{1}{3} \times 5 + \frac{1}{3} \times 1 + \frac{1}{3} \times 4 = 3 + \frac{1}{3}$$

from a game played in this way. This is illustrated in Figure 3.4 where the black spade shows the expected outcome of this mode of play. Auman called it a "correlated equilibrium". Indeed we can now mix these equilibria and still keep the correlated equilibrium property, as indicated by the doted line on Figure 3.4; also we can construct an expanded game in extensive form for which the correlated equilibrium constructed as above defines a Nash equilibrium (see Exercise 3.6).

In the above example we have seen that, by expanding the game via the adjunction of a first stage where "Nature" plays and gives information to the players, a new class of equilibria can be reached that dominate, in the outcome space, some of the original Nash-equilibria. If the random device gives an information which is common to all players, then it permits a mixing of the different pure strategy Nash equilibria and the outcome is in the convex hull of the Nash equilibrium outcomes. If the random device gives an information which may be different from one player to the other, then the correlated equilibrium can have an outcome which lies out of the convex hull of Nash equilibrium outcomes.

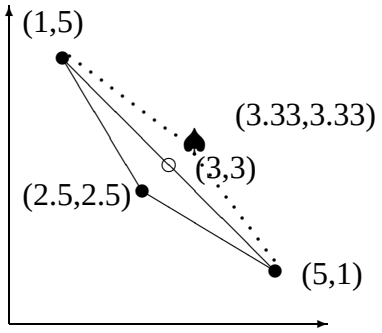


Figure 3.4: The dominating correlated equilibrium

3.6.2 A general definition of correlated equilibria

Let us give a general definition of a correlated equilibrium in an m-player normal form game. We shall actually give two definitions. The first one describes the construct of an *expanded game* with a random device distributing some pre-play information to the players. The second definition which is valid for m-matrix games is much simpler although equivalent.

Nash equilibrium in an expanded game

Assume that a game is described in normal form, with m players $j = 1, \ldots, m$, their respective strategy sets Γ_j and payoffs $V_j(\gamma_1, \ldots, \gamma_j, \ldots, \gamma_m)$. This will be called the *original normal form game*.

Assume the players may enter into a phase of *pre-play communication* during which they design a *correlation device* that will provide randomly a signal called *proposed mode of play*. Let $E = \{1, 2, \ldots, L\}$ be the finite set of the possible modes of play. The correlation device will propose with probability $\lambda(\ell)$ the mode of play ℓ. The device will then give the different players some information about the *proposed mode of play*. More precisely, let H_j be a class of subsets of E, called the *information structure* of player j. Player j, when the mode of play ℓ has been selected, receives an information denoted $h_j(\ell) \in H_j$. Now, we associate with each player j a *meta strategy* denoted $\tilde{\gamma}_j : H_j \to \Gamma_j$ that determines a strategy for the original normal form game, on the basis of the information received. All this construct, is summarized by the data $(E, \{\lambda(\ell)\}_{\ell \in E}, \{h_j(\ell) \in H_j\}_{j \in M, \ell \in E}, \{\tilde{\gamma}_j^* : H_j \to \Gamma_j\}_{j \in M})$ that defines an *expanded game*.

Definition 3.6.1 *The data* $(E, \{\lambda(\ell)\}_{\ell \in E}, \{h_j(\ell) \in H_j\}_{j \in M, \ell \in E}, \{\tilde{\gamma}_j^* : H_j \to \Gamma_j\}_{j \in M})$

defines a correlated equilibrium of the original normal form game if it is a Nash equilibrium for the expanded game i.e. if no player can improve his expected payoff by changing unilaterally his meta strategy, i.e. by playing $\tilde{\gamma}_j(h_j(\ell))$ instead of $\tilde{\gamma}_j^*(h_j(\ell))$ when he receives the signal $h_j(\ell) \in H_j$.

$$\sum_{\ell \in E} \lambda(\ell) V_j([\tilde{\gamma}_j^*(h_j(\ell)), \tilde{\gamma}_{M-j}^*(h_{M-j}(\ell))] \geq \sum_{\ell \in E} \lambda(\ell) V_j([\tilde{\gamma}_j(h_j(\ell)), \tilde{\gamma}_{M-j}^*(h_{M-j}(\ell))]. \tag{3.75}$$

An equivalent definition for m-matrix games

In the case of an m-matrix game, the definition given above can be replaced with the following one, which is much simpler

Definition 3.6.2 *A correlated equilibrium is a probability distribution $\pi(s)$ over the set of pure strategies $S = S_1 \times S_2 \ldots \times S_m$ such that, for every player j and any mapping $\delta_j : S_j \to S_j$ the following holds*

$$\sum_{s \in S} \pi(s) V_j([s_j, s_{M-j}]) \geq \sum_{s \in S} \pi(s) V_j([\delta(s_j), s_{M-j}]). \tag{3.76}$$

3.7 Bayesian equilibrium with incomplete information

Up to now we have considered only games where each player knows everything concerning the rules, the players types (*i.e.* their payoff functions, their strategy sets) etc... We were dealing with *games of complete information*. In this section we look at a particular class of *games with incomplete information* and we explore the class of so-called *Bayesian equilibria*.

3.7.1 Example of a game with unknown type for a player

In a game of incomplete information, some players do not know exactly what are the characteristics of other players. For example, in a two-player game, Player 2 may not know exactly what the payoff function of Player 1 is.

Example 3.7.1 *Consider the case where Player 1 could be of two types, called θ^1 and θ^2 respectively. We define below two matrix games, corresponding to the two possible*

types respectively:

θ^1	c_1	c_2
r_1	$0,-1$	$2,0$
r_2	$2,1$	$3,0$

$Game\ 1$

θ^2	c_1	c_2
r_1	$1.5,-1$	$3.5,0$
r_2	$2,1$	$3,0$

$Game\ 2$

If Player 1 is of type θ^1, then the bimatrix game 1 is played; if Player 1 is of type θ^2, then the bimatrix game 2 is played; the problem is that Player 2 does not know the type of Player 1.

3.7.2 Reformulation as a game with imperfect information

Harsanyi, in [Harsanyi, 1967-68], has proposed a transformation of a game of *incomplete information* into a game with *imperfect information*. For the example 3.7.1 this transformation introduces a preliminary *chance move*, played by *Nature* which decides randomly the type θ^i of Player 1. The probability distribution of each type, denoted p^1 and p^2 respectively, represents the beliefs of Player 2, given here in terms of prior probabilities, about facing a player of type θ^1 or θ^2. One assumes that Player 1 knows also these beliefs, the prior probabilities are thus *common knowledge*. The information structure[5] in the associated extensive game shown in Figure 3.5, indicates that Player 1 knows his type when deciding, whereas Player 2 does not observe the type neither, indeed in that game of simultaneous moves, the action chosen by Player 1. Call x_i (resp. $1-x_i$) the probability of choosing r_1 (resp. r_2) by Player 1 when he implements a mixed strategy, knowing that he is of type θ^i. Call y (resp. $1-y$) the probability of choosing c_1 (resp. c_2) by Player 2 when he implements a mixed strategy.

We can define the optimal response of Player 1 to the mixed strategy $(y, 1-y)$ of Player 2 by solving[6]

$$\max_{i=1,2} a_{i1}^{\theta^1} y + a_{i2}^{\theta^1}(1-y)$$

if the type is θ^1,

$$\max_{i=1,2} a_{i1}^{\theta^2} y + a_{i2}^{\theta^2}(1-y)$$

if the type is θ^2.

We can define the optimal response of Player 2 to the pair of mixed strategy $(x_i, 1-x_i), i=1,2$ of Player 1 by solving

$$\max_{j=1,2} p_1(x_1 b_{1j}^{\theta^1} + (1-x_1) b_{2j}^{\theta^1}) + p_2(x_2 b_{1j}^{\theta^2} + (1-x_2) b_{2j}^{\theta^2}).$$

[5]The dotted line in Figure 3.5 represents the information set of Player 2.

[6]We call $a_{ij}^{\theta^\ell}$ and $b_{ij}^{\theta^\ell}$ the payoffs of Player 1 and Player 2 respectively when the type is θ^ℓ.

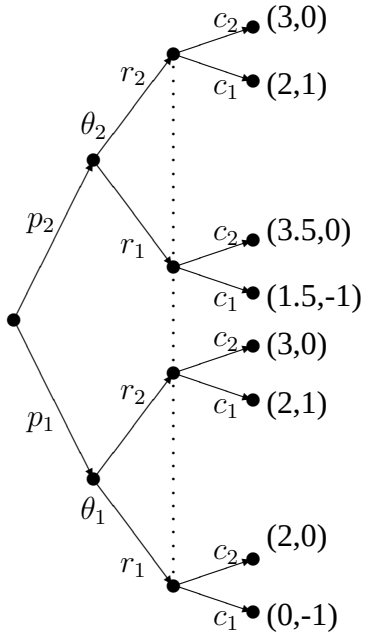


Figure 3.5: Extensive game formulation of the game of incomplete information

Let us rewrite these conditions with the data of the game illustrated in Figure 3.5. First consider the reaction function of Player 1

$$\begin{aligned} \theta_1 &\Rightarrow \max\{0y + 2(1-y), 2y + 3(1-y)\} \\ \theta_2 &\Rightarrow \max\{1.5y + 3.5(1-y), 2y + 3(1-y)\}. \end{aligned}$$

We draw in Figure 3.6 the lines corresponding to these comparisons between two linear functions. We observe that, if Player 1's type is θ^1, he will always choose r_2, whatever y, whereas, if his type is θ^2 he will choose r_1 if y is mall enough and switch to r_2 when $y > 0.5$. For $y = 0.5$, the best reply of Player 1 could be any mixed strategy $(x, 1-x)$.

Consider now the optimal reply of Player 2. We know that Player 1, if he is of type θ_1 xhooses always action r_2 *i.e.* $x_1 = 0$. So the best reply conditions for Player 2 can be written as follows

$$\max\{(p_1(-1) + (1-p_1)(x_2(-1) + (1-x_2)1)), (p_1(0) + (1-p_1)(x_2(0) + (1-x_2)0))\}$$

which boils down to

$$\max\{(1 - 2(1-p_1)x), 0)\}.$$

We conclude that Player 2 chooses action c_1 if $x_2 < \frac{1}{2(1-p_1)}$, action c_2 if $x_2 > \frac{1}{2(1-p_1)}$ and any mixed action with $y \in [0,1]$ if $x_2 = \frac{1}{2(1-p_1)}$. We conclude easily from these observations that the equilibria of this game are characterized as follows:

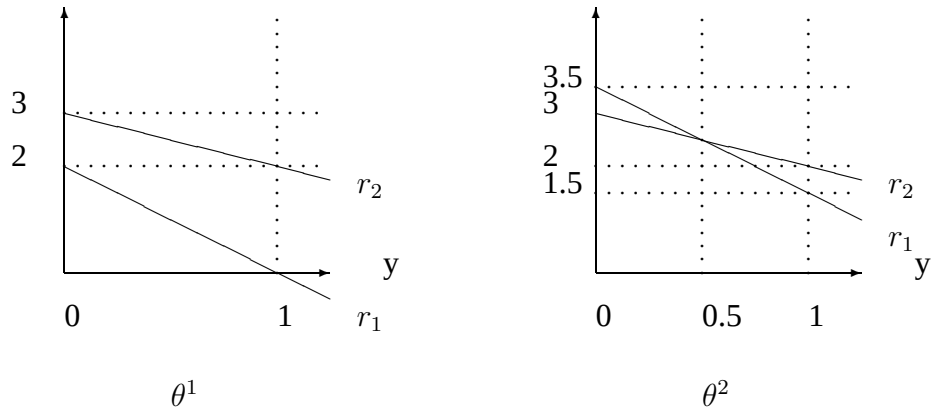


Figure 3.6: Optimal reaction of Player 1 to Player 2 mixed strategy $(y, 1-y)$

$x_1 \equiv 0$
if $p_1 \leq 0.5 \quad x_2 = 0, y = 1$ or $x_2 = 1, y = 0$ or $x_2 = \frac{1}{2(1-p_1)}, y = 0.5$
if $p_1 > 0.5 \quad x_2 = 0, y = 1.$

3.7.3 A general definition of Bayesian equilibria

We can generalize the analysis performed on the previous example and introduce the following definitions. Let M be a set of m players. Each player $j \in M$ may be of different possible types. Let Θ_j be the finite set of types for Player j. Whatever his type, Payer j has the same set of pure strategies $\mathcal{S}_j$. If $\theta = (\theta_1, \ldots . \theta_m) \in \Theta = \Theta_1 \times \ldots \times \Theta_m$ is a type specification for every player, then the normal form of the game is specified by the payoff functions

$$u_j(\theta; \cdot, \ldots, \cdot) : \mathcal{S}_1 \times \ldots \times \mathcal{S}_m \to \mathbf{R}, \quad j \in M. \tag{3.77}$$

A prior probability distribution $p(\theta_1, \ldots . \theta_m)$ on Θ is given as *common knowledge*. We assume that all the marginal probabilities are nonzero

$$p_j(\theta_j) > 0, \forall j \in M.$$

When Player j observes that he is of type $\theta_j \in \Theta_j$ he can construct his revised conditional probality distribution on the types θ_{M-j} of the other players through the Bayes formula

$$p(\theta_{M-j}|\theta_j) = \frac{p([\theta_j, \theta_{M-j}])}{p_j(\theta_j)}. \tag{3.78}$$

We can now introduce an *expanded game* where Nature draws randomly, according to the prior probability distribution $p(\cdot)$, a type vector $\theta \in \Theta$ for all players. Player-j can observe only his own type $\theta_j \in \Theta_j$. Then each player $j \in M$ picks a strategy in his own strategy set $\mathcal{S}_j$. The outcome is then defined by the payoff functions (3.77).

Definition 3.7.1 *In a game of incomplete information with m players having respective type sets Θ_j and pure strategy sets $\mathcal{S}_j$, $i = 1, \ldots, m$, a Bayesian equilibrium is a Nash equilibrium in the " expanded game " in which each player pure strategy γ_j is a map from Θ_j to $\mathcal{S}_j$.*

The expanded game can be described in its normal form as follows. Each player $j \in M$ has a strategy set Γ_j where a strategy is defined as a mapping $\gamma_j : \Theta_j \to \mathcal{S}_j$. Associated with a strategy profile $\gamma = (\gamma_1, \ldots, \gamma_m)$ the payoff to player j is given by

$$\begin{aligned} V_j(\gamma) &= \sum_{\theta_j \in \Theta_j} p_j(\theta_j) \sum_{\theta_{M-j} \in \Theta_{M-j}} p(\theta_{M-j}|\theta_j) \\ &\quad u_j([\theta_j, \theta_{M-j}]; \gamma_1(\theta_1), \ldots, \gamma_j(\theta_j), \ldots, \gamma_m(\theta_m)). \end{aligned} \tag{3.79}$$

As usual, a Nash equilibrium is a strategy profile $\gamma^* = (\gamma_1^*, \ldots, \gamma_m^*) \in \Gamma = \Gamma_1 \times \ldots \times \Gamma_m$ such that

$$V_j(\gamma^*) \geq V_j(\gamma_j, \gamma_{M-j}^*) \quad \forall \gamma_j \in \Gamma_j. \tag{3.80}$$

It is easy to see that, since each $p_j(\theta_j)$ is positive, the equilibrium conditions (3.80) lead to the following conditions

$$\gamma_j^*(\theta_j) = \operatorname{argmax}_{s_j \in \mathcal{S}_j} \sum_{\theta_{M-j} \in \Theta_{M-j}} p(\theta_{M-j}|\theta_j)\, u_j([\theta_j, \theta_{M-j}]; [s_j, \gamma_{M-j}^*(\theta_{M-j})]) \quad j \in M. \tag{3.81}$$

Remark 3.7.1 *Since the sets Θ_j and $\mathcal{S}_j$ are finite, the set Γ_j of mappings from Θ_j to $\mathcal{S}_j$ is also finite. Therefore the expanded game is an m-matrix game and, by Nash theorem there exists at least one mixed strategy equilibrium.*

3.8 Appendix on Kakutani Fixed-point theorem

Definition 3.8.1 *Let $\Phi : \boldsymbol{R}^m \mapsto 2^{\boldsymbol{R}^n}$ be a point to set mapping. We say that this mapping is upper semicontinuous if, whenever the sequence $\{x_k\}_{k=1,2,\ldots}$ converges in $\boldsymbol{R}^m$ toward x^0 then any accumulation point y^0 of the sequence $\{y_k\}_{k=1,2,\ldots}$ in $\boldsymbol{R}^n$, where $y_k \in \Phi(x_k)$, $k = 1, 2, \ldots$, is such that $y^0 \in \Phi(x_0)$.*

Theorem 3.8.1 *Let $\Phi : A \mapsto 2^A$ be a point to set upper semicontinuous mapping, where A is a compact subset[7] of $\boldsymbol{R}^m$. Then there exists a fixed-point for Φ. That is, there is $x^* \in \Phi(x^*)$ for some $x^* \in A$.*

[7] i.e. a closed and bounded subset.

3.9 exercises

Exercise 3.1: Consider the matrix game.

$$\begin{bmatrix} 3 & 1 & 8 \\ 4 & 10 & 0 \end{bmatrix}$$

Assume that the players are in the simultaneous move information structure. Assume that the players try to guess and counterguess the optimal behavior of the opponent in order to determine their optimal strategy choice. Show that this leads to an unstable process.

Exercise 3.2: Find the value and the saddle point mixed-strategies for the above matrix game.

Exercise 3.3: Define the quadratic programming problem that will find a Nash equilibrium for the bimatrix game

$$\begin{bmatrix} (52,50)^* & (44,44) & (44,41) \\ (42,42) & (46,49)^* & (39,43) \end{bmatrix}.$$

Verify that the entries marked with a $*$ correspond to a solution of the associated quadratic programming problem.

Exercise 3.4: Do the same as above but using now the complementarity problem formulation.

Exercise 3.5: Consider the two player game where the strategy sets are the intervals $U_1 = [0, 100]$ and $U_2 = [0, 100]$ respectively, and the payoffs $\psi_1(\mathbf{u}) = 25u_1 + 15u_1u_2 - 4u_1^2$ and $\psi_2(\mathbf{u}) = 100u_2 - 50u_1 - u_1u_2 - 2u_2^2$ respectively. Define the best reply mapping. Find an equilibrium point. Is it unique?

Exercise 3.6: Consider the two player game where the strategy sets are the intervals $U_1 = [10, 20]$ and $U_2 = [0, 15]$ respectively, and the payoffs $\psi_1(\mathbf{u}) = 40u_1 + 5u_1u_2 - 2u_1^2$ and $\psi_2(\mathbf{u}) = 50u_2 - 3u_1u_2 - 2u_2^2$ respectively. Define the best reply mapping. Find an equilibrium point. Is it unique?

Exercise 3.7: Consider an oligopoly model. Show that the assumptions 3.5.1 define a concave game in the sense of Rosen.

Exercise 3.6: In example 3.7.1 a correlated equilibrium has been constructed for the game

	c_1	c_2
r_1	$5, 1$	$0, 0$
r_2	$4, 4$	$1, 5$

.

Find the associated extensive form game for which the proposed correlated equilibrium corresponds to a Nash equilibrium.

Part II

Repeated and sequential Games

Chapter 4

Repeated games and memory strategies

In this chapter we begin our analysis of "dynamic games". To say it in an imprecise but intuitive way, the dynamic structure of a game expresses a change over time of the conditions under which the game is played. In a repeated game this structural change over time is due to the accumulation of information about the "history" of the game. As time unfolds the information at the disposal of the players changes and, since strategies transform this information into actions, the players' strategic choices are affected. If a game is repeated twice, *i.e.* the same game is played in two successive stages and a reward is obtained at each stage, one can easily see that, at the second stage, the players can decide on the basis of the outcome of the first stage. The situation becomes more and more complex as the number of stages increases, since the players can base their decisions on *histories* represented by sequences of actions and outcomes observed over increasing numbers of stages. We may also consider games that are played over an infinite number of stages. These infinitely repeated games are particularly interesting, since, at each stage, there is always an infinite number of remaining stages over which the game will be played and, therefore, there will not be the so-called "end of horizon effect" where the fact that the game is ending up affects the players behavior. Indeed, the evaluation of strategies will have to be based on the comparison of infinite streams of rewards and we shall explore different ways to do that. In summary, the important concept to understand here is that, even if it is the "same game" which is repeated over a number of periods, the global "repeated game" becomes a fully dynamic system with a much more complex strategic structure than the one-stage game.

In this chapter we shall consider repeated bimatrix games and repeated concave m-player games. We shall explore, in particular, the class of equilibria when the number of periods over which the game is repeated (the *horizon*) becomes infinite. We will show how the class of equilibria in repeated games can be expended very much if one

allows memory in the information structure. The class of *memory strategies* allows the players to incorporate *threats* in their strategic choices. A threat is meant to deter the opponents to enter into a detrimental behavior. In order to be effective a threat must be *credible*. We shall explore the credibility issue and show how the subgame perfectness property of equilibria introduces a refinement criterion that make these threats credible.

Notice that, at a given stage, the players actions have no direct influence on the normal form description of the games that will be played in the forthcoming periods. A more complex dynamic structure would obtain if one assumes that the players actions at one stage influence the type of game to be played in the next stage. This is the case for the class of *Markov* or *sequential* games and a fortiori for the class of *differential games*, where the time flows continuously. These dynamic games will be discussed in subsequent chapters of this volume.

4.1 Repeating a game in normal form

Repeated games have been also called *games in semi-extensive form* in [Friedman 1977]. Indeed this corresponds to an explicit dynamic structure, as in the extensive form, but with a game in normal form defined at each period of time. In the repeated game, the global strategy becomes a way to define the choice of a stage strategy at each period, on the basis of the knowledge accumulated by the players on the "history" of the game. Indeed the repetition of the same game in normal form permits the players to adapt their strategies to the observed history of the game. In particular, they have the opportunity to implement a class of strategies that incorporate threats.

The payoff associated with a repeated game is usually defined as the sum of the payoffs obtained at each period. However, when the number of periods to tend to ∞, a total payoff that implies an infinite sum may not converge to a finite value. There are several ways of dealing with the comparison of infinite streams of payoffs. We shall discover some of them in the subsequent examples.

4.1.1 Repeated bimatrix games

Figure 4.1 describes a repeated bimatrix game structure. The same matrix game is played repeatedly, over a number T of *periods* or *stages* that represent the passing of time. In the one-stage game, where Player 1 has p pure strategies and Player 2 q pure strategies and the one-stage payoff pair associated with the startegy pair (k, ℓ) is given by $(\alpha^j_{k\ell})_{j=1,2}$. The payoffs are accumulated over time. As the players may recall the past history of the game, the extensive form of those repeated games is quite complex.

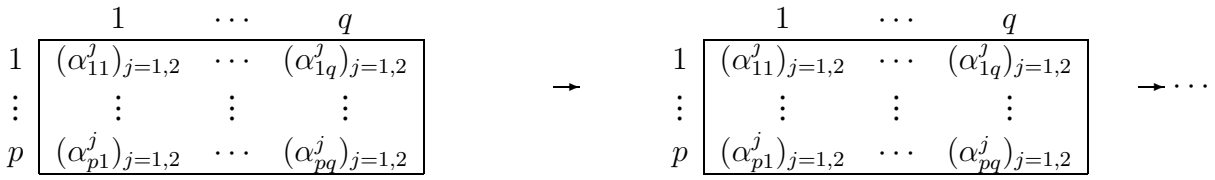


Figure 4.1: A repeated game

The game is defined in a *semi-extensive form* since at each period a normal form (*i.e.* an already aggregated description) defines the consequences of the strategic choices in a one-stage game. Even if the same game seems to be played at each period (it is similar to the others in its normal form description), in fact, the possibility to use the history of the game as a source of information to adapt the strategic choice at each stage makes a repeated game a fully dynamic "object".

We shall use this structure to prove an interesting result, called the "folk theorem", which shows that we can build an infinitely repeated bimatrix game, played with the help of finite automata, where the class of equilibria permits the approximation of any individually rational outcome of the one-stage bimatrix game.

4.1.2 Repeated concave games

Consider now the class of dynamic games where one repeats a concave game à la Rosen. Let $t \in \{0, 1, \ldots, T-1\}$ be the time index. At each time period t the players enter into a noncooperative game defined by

$$(M; U_j, \psi_j(\cdot), j \in M)$$

where as indicated in section 3.4, $M = \{1, 2, \ldots, m\}$ is the set of players, U_j is a compact convex set describing the actions available to Player j at each period and $\psi_j(\mathbf{u}(t))$ is the payoff to player j when the action vector $\mathbf{u}(t) = (u_1(t), \ldots, u_m(t)) \in \mathbf{U} = U_1 \times U_2 \times \cdots \times U_m$ is chosen at period t by the m players. This function is assumed to be concave in u_j.

We shall denote $\tilde{\mathbf{u}} = (\mathbf{u}(t) : t = 0, 1 \ldots, T-1)$ the actions sequence over the sequence of T periods. The total payoff of player j over the T-horizon is then defined by

$$V_j^T(\tilde{\mathbf{u}}) = \sum_{t=0}^{T-1} \psi_j(\mathbf{u}(t)). \tag{4.1}$$

Open-loop information structure

At period t, each player j knows only the current time t and what he has played in periods $\tau = 0, 1, \ldots, t-1$. He does not observe what the other players do. We implicitly assume that the players cannot use the rewards they receive at each period to *infer* through ***e.g.*** an appropriate filtering the actions of the other players. The *open-loop* information structure actually eliminates almost every aspect of the dynamic structure of the repeated game context.

Closed-loop information structure

At period t, each player j knows not only the current time t and what he has played in periods $\tau = 0, 1, \ldots, t-1$, but also what the other players have done at previous period. We call *history of the game at period* τ the sequence

$$h(\tau) = (\mathbf{u}(t) : t = 0, 1 \ldots, \tau - 1).$$

A closed-loop strategy for Player j is defined as a sequence of mappings

$$\gamma_j^\tau : h(\tau) \mapsto U_j.$$

Due to the repetition over time of the same game, each player j can adjust his choice of one-stage action u_j to the history of the game. This permits the consideration of *memory strategies* where some threats are included in the announced strategies. For example, a player can declare that some particular histories would "trigger" a *retaliation* from his part. The description of a *trigger strategy* will therefore include

- a nominal "mood" of play that contributes to the expected or desired outcomes
- a retaliation "mood" of play that is used as a threat
- the set of histories that trigger a switch from the *nominal mood of play* to the *retaliation mood of play*.

Many authors have studied equilibria obtained in repeated games through the use of *trigger strategies* (see for example [Radner, 1980], [Radner, 1981], [Friedman 1986]).

Infinite horizon games

If the game is repeated over an infinite number of periods then payoffs are represented by infinite streams of rewards $\psi_j(\mathbf{u}(t)), \quad t = 1, 2, \ldots, \infty$.

Discounted sums: We have assumed that the actions sets U_j are compact, therefore, since the functions $\psi_j(\cdot)$ are continuous, the one-stage rewards $\psi_j(\mathbf{u}(t))$ are uniformly bounded. Hence, the discounted payoffs

$$V_j(\tilde{\mathbf{u}}) = \sum_{t=0}^{\infty} \beta_j^t \psi_j(\mathbf{u}(t)), \tag{4.2}$$

where $\beta_j \in [0, 1)$ is the discount factor of Player j, are well defined. Therefore, if the players discount time, we can compare strategies on the basis of the discounted sums of the rewards they generate for each player.

Average rewards per period: If the players do not discount time the comparison of strategies imply the consideration of infinite streams of rewards that sum up to infinity. A way to circumvent the difficulty is to consider a *limit average reward per period* in the following way

$$g_j(\tilde{\mathbf{u}}) = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \psi_j(\mathbf{u}(t)). \tag{4.3}$$

To link this evaluation of infinite streams of rewards with the previous one based on discounted sums one may define the *equivalent discounted constant reward*

$$g_j^{\beta_j}(\tilde{\mathbf{u}}) = (1 - \beta_j) \sum_{t=0}^{\infty} \beta_j^t \psi_j(\mathbf{u}(t)). \tag{4.4}$$

One expects that, for well behaved games, the following limiting property holds true

$$\lim_{\beta_j \to 1} g_j^{\beta_j}(\tilde{\mathbf{u}}) = g_j(\tilde{\mathbf{u}}). \tag{4.5}$$

In infinite horizon repeated games each player has always *enough time to retaliate* . He will always have the possibility to implement an announced threat, if the opponents are not playing according to some *nominal sequence* corresponding to an agreement. We shall see that the equilibria, based on the use of *threats* constitute a very large class. Therefore the set of strategies with the *memory* information structure is much more encompassing than the *open-loop* one.

Existence of a "trivial" dynamic equilibrium

We know, by Rosen's existence theorem, that a concave "static game"

$$(M; U_j, \psi_j(\cdot), j \in M)$$

admits an equilibrium $\mathbf{u}^*$. It should be clear (see exercise 4.1) that the sequence $\tilde{\mathbf{u}}^* = (\mathbf{u}(t) \equiv \mathbf{u}^* : t = 0, 1 \ldots, T-1)$ is also an equilibrium for the repeated game in both the open-loop and closed-loop (memory) information structure.

Assume the one period game is such that a unique equilibrium exists. The open-loop repeated game will also have a unique equilibrium, whereas the closed-loop repeated game will still have a plethora of equilibria, as we shall see it shortly.

4.2 Folk theorem

In this section we present the celebrated *Folk theorem*. This name is due to the fact that there is no well defined author of the first version of it. The idea is that an infinitely repeated game should permit the players to design equilibria, supported by threats, with outcomes being Pareto efficient. The strategies will be based on *memory*, each player remembering the past moves of game. Indeed the infinite horizon could give raise to an infinite storage of information, so there is a question of designing strategies with the help of mechanisms that would exploit the history of the game but with a finite information storage capacity. Finite automata provide such systems.

4.2.1 Repeated games played by automata

A *finite automaton* is a logical system that, when stimulated by an *input*, generates an *output*. For example an automaton could be used by a player to determine the stage-game action he takes given the information he has received. The automaton is finite if the length of the logical input (a stream of $0-1$ bits) it can store is finite. So in our repeated game context, a finite automaton will not permit a player to determine his stage-game action upon an infinite memory of what has happened before. To describe a finite automaton one needs the following

- A list of all possible stored input configurations; this list is finite and each element is called a *state*. One element of this list is the *initial state*.
- An output function determines the action taken by the automaton, given its current state.
- A transition function tells the automaton how to move from one state to another after it has received a new input element.

In our paradigm of a repeated game played by automata, the input element for the automaton a of Player 1 is the stage-game action of Player 2 in the preceding stage, and symmetrically for Player 2. In this way, each player can choose a way to process the information he receives in the course of the repeated plays, in order to decide the next action.

An important aspect of finite automata is that, when they are used to play a repeated game, they necessarily generate cycles in the successive states and hence in the actions taken in the successive stage-games.

Let G be a finite two player game, i.e. a bimatrix game, which defines the one-stage game. U and V are the sets of pure strategies in the game G for Player 1 and 2 respectively. Denote by $g_j(u, v)$ the payoff to Player j in the one-stage game when the pure strategy pair $(u, v) \in U \times V$ has been selected by the two players.

The game is repeated indefinitely. Although the set of strategies for an infinitely repeated game is enormously rich, we shall restrict the strategy choices of the players to the class of finite automata. A pure strategy for Player 1 is an automaton $a \in A$ which has for input the actions $u \in U$ of Player 2. Symmetrically a pure strategy for Player 2 is an automaton $b \in B$ which has for input the actions $v \in V$ of Player 1.

We associate with the pair of finite automata $(a, b) \in A \times B$ a pair of average payoffs per stage defined as follows

$$\tilde{g}_j(a, b) = \frac{1}{N} \sum_{n=1}^{N} g_j(u_n, v_n) \quad j = 1, 2, \tag{4.6}$$

where N is the length of the cycle associated with the pair of automata (a, b) and (u_n, v_n) is the action pair at the n-th stage in this cycle. One notices that the expression (4.6) is also the limite average reward per period due to the cycling behavior of the two automata.

We call $\mathcal{G}$ the game defined by the strategy sets A and B and the payoff functions (4.6).

4.2.2 Minimax point

Assume Player 1 wants to threaten Player 2. An effective threat would be to define his action in a one-stage game through the solution of

$$\bar{m}_2 = \min_u \max_v g_2(u, v).$$

Similarly if Player 2 wants to threaten Player 2, an effective threat would be to define his action in a one-stage game through the solution of

$$\bar{m}_1 = \min_v \max_u g_1(u, v).$$

We call *minimax point* the pair $\bar{m} = (\bar{m}_1, \bar{m}_2)$ in $\mathbf{R}^2$.

4.2.3 Set of outcomes dominating the minimax point

Theorem 4.2.1 *Let $\mathcal{P}_{\bar{m}}$ be the set of outcomes in a matrix game that dominate the minimax point $\bar{m}$. Then the outcomes corresponding to Nash equilibria in pure strategies*[1] *of the game $\mathcal{G}$ are dense in $\mathcal{P}_{\bar{m}}$.*

Proof: Let $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K$ be the pairs of payoffs that appear in the bimatrix game when it is written in the vector form, that is

$$\mathbf{g}_k = (g_1(u_k, v_k), g_2(u_k, v_k))$$

where (u_k, v_k) is a possible pure strategy pair. Let $q_1, q_2, \ldots, q_K$ be nonnegative rational numbers that sum to 1. Then the convex combination

$$\mathbf{g}^* = q_1\mathbf{g}_1 + q_2\mathbf{g}_2, \ldots, q_K\mathbf{g}_K$$

is an element of $\mathcal{P}$ if $g_1^* \geq \bar{m}_1$ and $g_2^* \geq \bar{m}_2$. The set of these $\mathbf{g}$ is dense in $\mathcal{P}$. Each q_k being a rational number can be written as a ratio of two integers, that is a fraction. It is possible to write the K fractions with a common denominator, hence

$$q_k = \frac{n_k}{N},$$

with

$$n_1 + n_2 + \ldots + n_K = N$$

since the q_k's must sum to 1.

We construct two automata a^* and b^*, such that, together, they play (u_1, v_1) during n_1 stages, (u_2, v_2) during n_2 stages, etc. They complete the sequence by playing (u_K, v_K) during n_K stages and the N-stage cycle begins again. The limit average per period reward of Player j in the $\mathcal{G}$ game played by these automata is given by

$$\tilde{g}_j(a^*, b^*) = \frac{1}{N}\sum_{k=1}^{K} n_k g_j(u_k, v_k) = \sum_{k=1}^{K} q_k g_j(u_k, v_k) = g_j^*.$$

Therefore these two automata will achieve the payoff pair $\mathbf{g}$. Now we refine the structure of these automata as follows:

- Let $\tilde{u}(n)$ and $\tilde{v}(n)$ be the pure strategy that Player 1 and Player 2 respectively are supposed to play at stage n, according to the cycle defined above.

[1] A pure strategy in the game $\mathcal{G}$ is a deterministic choice of an automaton; this automaton will process information coming from the repeated use of mixed strategies in the repeated game.

- The automaton a^* plays $\tilde{u}(n+1)$ at stage $n+1$ if automaton b has played $\tilde{v}(n)$ at stage n; otherwise it plays $\bar{u}$, the strategy that is minimaxing Player 2's one shot payoff and this strategy will be played forever.
- Symmetrically, the automaton b^* plays $\tilde{v}(n+1)$ at stage $n+1$ if automaton a has played $\tilde{u}(n)$ at stage n; otherwise it plays $\bar{v}$, the strategy that is minimaxing Player 1's one shot payoff and this strategy will be played forever.

It should be clear that the two automata a^* and b^* defined as above constitute a Nash equilibrium for the repeated game $\mathcal{G}$ whenever $\mathbf{g}$ is dominating $\bar{m}$, i.e. $\mathbf{g} \in \mathcal{P}_{\bar{m}}$. Indeed, if Player 1 changes unilaterally his automaton to a then the automaton b^* will select for all periods, except a finite number, the minimax action $\bar{v}$ and, as a consequence, the limit average per period reward obtained by Player 1 is

$$\tilde{g}_1(a, b^*) \leq \bar{m}_1 \leq \tilde{g}_1(a^*, b^*).$$

Similarly for Player 2, a unilateal change to $b \in B$ leads to a payoff

$$\tilde{g}_2(a^*, b) \leq \bar{m}_2 \leq \tilde{g}_2(a^*, b^*).$$

Q.E.D.

Remark 4.2.1 *This result is known as the Folk theorem, because it has always been in the "folklore" of game theory without knowing to whom it should be attributed.*

The class of equilibria is more restricted if one imposes a condition that the threats used by the players in their announced strategies have to be *credible*. This is the issue discussed in one of the following sections.

4.3 Collusive equilibrium in a repeated Cournot game

Consider a Cournot oligopoly game repeated over an infinite sequence of periods. Let $t = 1, \ldots, \infty$ be the sequence of time periods. Let $\mathbf{q}(t) \in \mathbf{R}^m$ be the production vector chosen by the m firms at period t and $\psi_j(\mathbf{q}(t))$ the payoff to player j at period t.

We denote

$$\tilde{\mathbf{q}} = \mathbf{q}(1), \ldots, \mathbf{q}(t), \ldots$$

the infinite sequence of production decisions of the m firms. Over the infinite time horizon the rewards of the players are defined as

$$\tilde{V}_j(\tilde{\mathbf{q}}) = \sum_{t=0}^{\infty} \beta_j^t \psi_j(\mathbf{q}(t))$$

where $\beta_j \in [0,1)$ is a discount factor used by Player j.

We assume that at each period t all players have the same information $h(t)$ which is the history of passed productions $h(t) = (\mathbf{q}(0), \mathbf{q}(1), \ldots, \mathbf{q}(t-1))$.

We consider, for the one-stage game, two possible outcomes,

(i) the Cournot equilibrium, supposed to be uniquely defined by $\mathbf{q}_c = (q_j^c)_{j=1,\ldots,m}$ and

(ii) a Pareto outcome resulting from the production levels $\mathbf{q}^* = (q_j^*)_{j=1,\ldots,m}$ and such that $\psi_j(\mathbf{q})^* \geq \psi_j(\mathbf{q^c})$, $j = 1, \ldots, m$. We say that this Pareto outcome is a collusive outcome dominating the Cournot equilibrium.

A Cournot equilibrium for the repeated game is defined by the strategy consisting for each Player j to choose repeatedly the production levels $q_j(t) \equiv q_j^c$, $j = 1, 2$. However there are many other equilibria that can be obtained through the use of memory strategies.

Let us define

$$\tilde{V}_j^* = \sum_{t=0}^{\infty} \beta_j^t \psi_j(\mathbf{q}^*) \tag{4.7}$$

$$\Phi_j(\mathbf{q}^*) = \max_{q_j} \psi_j([\mathbf{q}^{*(-j)}, q_j]). \tag{4.8}$$

The following has been shown in [Friedman 1977]

Lemma 4.3.1 *There exists an equilibrium for the repeated Cournot game that yields the payoffs $\tilde{V}_j^*$ if the following inequality holds*

$$\beta_j \geq \frac{\Phi_j(\mathbf{q}^*) - \psi_j(\mathbf{q}^*)}{\Phi_j(\mathbf{q}^*) - \psi_j(\mathbf{q}^c)}. \tag{4.9}$$

This equilibrium is reached through the use of a so-called trigger strategy defined as follows

$$\left.\begin{array}{ll} \textit{if} & (\mathbf{q}(0), \mathbf{q}(1), \ldots, \mathbf{q}(t-1)) = (\mathbf{q}^*, \mathbf{q}^*, \ldots, \mathbf{q}^*) \textit{ then } q_j(t) = q_j^* \\ \textit{otherwise} & q_j(t) = q_j^c. \end{array}\right\} \tag{4.10}$$

Proof: If the players play according to (4.10) the payoff they obtain is given by

$$\tilde{V}_j^* = \sum_{t=0}^{\infty} \beta_j^t \psi_j(\mathbf{q}^*) = \frac{1}{1-\beta_j} \psi_j(\mathbf{q}^*).$$

Assume Player j decides to deviate unilaterally at period 0. He knows that his deviation will be detected at period 1 and that, thereafter the Cournot equilibrium production level $\mathbf{q}^c$ will be played. So, the best he can expect from this deviation is the following payoff which combines the maximal reward he can get in period 0 when the other firms play q_k^* with the discounted value of an infinite stream of Cournot outcomes.

$$\tilde{V}_j(\tilde{\mathbf{q}}) = \Phi_j(\mathbf{q}^*) + \frac{\beta_j}{1-\beta_j}\psi_j(\mathbf{q}^c).$$

This unilateral deviation is not profitable if the following inequality holds

$$\psi_j(\mathbf{q}^*) - \Phi_j(\mathbf{q}^*) + \frac{\beta_j}{1-\beta_j}(\psi_j(\mathbf{q}^*) - \psi_j(\mathbf{q}^c)) \geq 0,$$

which can be rewritten as

$$(1-\beta_j)(\psi_j(\mathbf{q}^*) - \Phi_j(\mathbf{q}^*)) + \beta_j(\psi_j(\mathbf{q}^*) - \psi_j(\mathbf{q}^c)) \geq 0,$$

from which we get

$$\psi_j(\mathbf{q}^*) - \Phi_j(\mathbf{q}^*) - \beta_j(\psi_j(\mathbf{q}^c) - \Phi_j(\mathbf{q}^*)) \geq 0.$$

and, finally the condition (4.9)

$$\beta_j \geq \frac{\psi_j(\mathbf{q}^*) - \Phi_j(\mathbf{q}^*)}{\Phi_j(\mathbf{q}^*) - \psi_j(\mathbf{q}^c)}.$$

The same reasonning holds if the deviation occurs at any other period t. **Q.E.D.**

Remark 4.3.1 *According to this strategy the* $\mathbf{q}^*$ *production levels correspond to a cooperative mood of play while the* $\mathbf{q}^c$ *correspond to a* punitive *mood of play. The punitive mood is a* threat. *As the punitive mood of play consists to play the Cournot solution which is indeed an equilibrium, it is a credible threat. Therefore the players will always choose to play in accordance with the Pareto solution* $\mathbf{q}^*$ *and thus the equilibrium defines a nondominated outcome.*

It is important to notice the difference between the threats used in the folk theorem and the ones used in this repeated Cournot game. Here the threats are constituting themselves an equilibrium for the dynamic game. We say that the memory (trigger strategy) equilibrium is *subgame perfect* in the sense of Selten [Selten, 1975].

4.3.1 Finite vs infinite horizon

There is an important difference between finitely and infinitely repeated games. If the same game is repeated only a finite number of times then, in a subgame perfect equilibrium, a single game equilibrium will be repeated at each period. This is due to the

impossibility to retaliate at the last period. Then this situation is translated backward in time until the initial period. When the game is repeated over an infinite time horizon there is always enough time to retaliate and possibly resume cooperation. For example, let us consider a repeated Cournot game, without discounting where the players' payoffs are determined by the long term average of the one sage rewards. If the players have *perfect information with delay* they can observe without errors the moves selected by their opponents in previous stages. There might be a delay of several stages before this observation is made. In that case it is easy to show that, to any cooperative solution dominating a static Nash equilibrium corresponds a subgame perfect sequential equilibrium for the repeated game, based on the use of *trigger strategies*. These strategies use as a threat a switch to the noncooperative solution for all remainig periods, as soon as a player has been observed not to play according to the cooperative solution. Since the long term average only depends on the infinitely repeated one stage rewards, it is clear that the threat gives raise to the Cournot equilibrium payoff in the long run. This is the subgame perfect version of the *Folk theorem* [Friedman 1986].

4.3.2 A repeated stochastic Cournot game with discounting and imperfect information

We give now an example of stochastic repeated game for wich the construction of equilibria based on the use of threats is not trivial. This is the stochastic repeated game, based on the Cournot model with imperfect information, (see [Green & Porter, 1985], [Porter, 1983]).

Let $X(t) = \sum_{j=1}^{m} x_j(t)$ be the total supply of m firms at time t and $\theta(t)$ be a sequence of i.i.d[2] random variables affecting these prices. Each firm j chooses its supplies $x_j(t) \geq 0$, in a desire to maximize its total expected discounted profit

$$V_j(x_1(\cdot), \ldots, x_j(\cdot), \ldots, x_m(\cdot)) = \mathrm{E}_\theta \sum_{t=0}^{\infty} \beta^t [D(X(t), \theta(t)) x_j(t) - c_j(x_j(t))], \quad j = 1, \ldots, m,$$

where $\beta < 1$ is the discount factor. The assumed information structure does not allow each player to observe the opponent actions used in the past. Only the sequence of past prices, corrupted by the random noise, is available. Therefore the players cannot monitor without error the past actions of their opponent(s). This impossibility to detect without error a breach of cooperation increases considerably the difficulty of building equilibria based on the use of credible threats. In [Green & Porter, 1985], [Porter, 1983] it is shown that there exist subgame perfect equilibria, based on the use of *trigger strategies*, which are dominating the repeated Cournot-Nash equilibrium. This construct uses the concept of Markov game and will be further discussed in Part 3

[2]Independent and identically distributed.

of these Notes. It will be shown that these equilibria are in fact closely related to the so-called *correlated* and/or *communication equilibria* discussed at the end of Part 2.

4.4 Exercises

Exercise 5.1. Prove that the repeated one-stage equilibrium is an equilibrium for the repeated game with additive payoffs and a finite number of periods. Extend the result to infinitely repeated games.

Exercise 5.2. Show that condition (4.9) holds if $\beta_j = 1$. Adapt Lemma 4.3.1 to the case where the game is evaluated through the long term average reward criterion.

Chapter 5

Shapley's Zero Sum Markov Game

5.1 Process and rewards dynamics

The concept of *Markov game* has been introduced, in a zero-sum game framework, in [Shapley. 1953]. The structure of this game can also be described as a *controlled Markov chain* with *two competing agents*.

Let $S = \{1, 2, \ldots, n\}$ be the set of possible states of a discrete time stochastic process $\{x(t) : t = 0, 1, \ldots\}$. Let $U_j = \{1, 2, \ldots, \lambda_j\}$, $j = 1, 2$, be the finite *action sets* of two players. The process dynamics is described by the *transition probabilities*

$$p_{s,s'}(\mathbf{u}) = \mathrm{P}[x(t+1) = s' | x(t) = s,\ \mathbf{u}], \quad s, s' \in S, \quad \mathbf{u} \in U_1 \times U_2,$$

which satisfy, for all $\mathbf{u} \in U_1 \times U_2$

$$\begin{aligned} p_{s,s'}(\mathbf{u}) &\geq 0 \\ \sum_{s' \in S} p_{s,s'}(\mathbf{u}) &= 1, \quad s \in S. \end{aligned}$$

As the transition probabilities depend on the players actions we speak of a *controlled Markov chain.* A transition reward function

$$r(s, \mathbf{u}), \quad s \in S, \quad \mathbf{u} \in U_1 \times U_2,$$

defines the gain of player 1 when the process is in state s and the two players take the action pair $\mathbf{u}$. Player 2's reward is given by $-r(s, \mathbf{u})$, since the game is assumed to be *zero-sum*.

5.2 Information structure and strategies

5.2.1 The extensive form of the game

The game defined above corresponds to a game in extensive form with an infinite number of moves. We assume an information structure where the players choose *sequentially* and *simultaneously*, with *perfect recall*, their actions. This is illustrated in figure 5.1

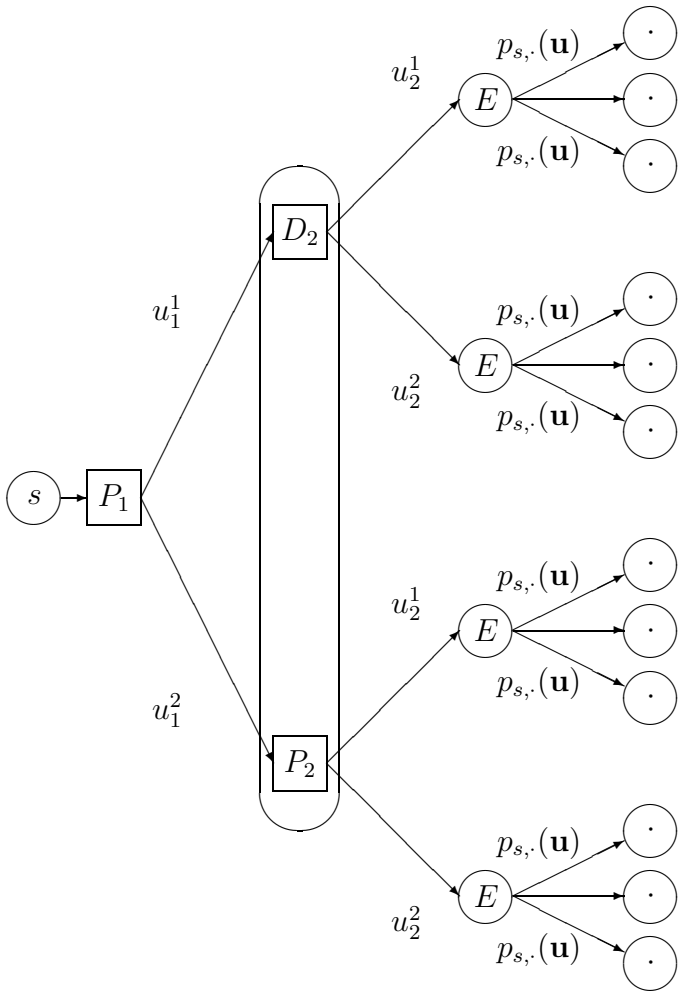


Figure 5.1: A Markov game in extensive form

5.2.2 Strategies

A strategy is a device which transforms the information into action. Since the players may recall the past observations, the general form of a strategy can be quite complex.

Markov strategies: These strategies are also called *feedback strategies*. The assumed information structure allows the use of strategies defined as mappings

$$\gamma_j : S \mapsto P(U_j), \quad j = 1, 2,$$

where $P(U_j)$ is the class of probability distributions over the action set U_j. Since the strategies are based only on the information conveyed by the current state of the x-process, they are called *Markov strategies*.

When the two players have chosen their respective strategies the state process becomes a Markov chain with transition probabilities

$$\mathrm{P}_{s,s'}^{\gamma_1,\gamma_2} = \sum_{k=1}^{\lambda_1} \sum_{\ell=1}^{\lambda_2} \mu_k^{\gamma_1(s)} \nu_\ell^{\gamma_2(s)} p_{s,s'}(u_1^k, u_2^\ell),$$

where we have denoted $\mu_k^{\gamma_1(s)}, \quad k = 1, \ldots, \lambda_1$, and $\nu_\ell^{\gamma_2(s)}, \quad \ell = 1, \ldots, \lambda_2$ the probability distributions induced on U_1 and U_2 by $\gamma_1(s)$ and $\gamma_2(s)$ respectively. In order to formulate the normal form of the game, it remains to define the payoffs associated with the admissible strategies of the twp players. Let $\beta \in [0, 1)$ be a given discount factor. Player 1's payoff, when the game starts in state s_0, is defined as the discounted sum over the infinite time horizon of the expected transitions rewards, i.e.

$$V(s_0; \gamma_1, \gamma_2) = \mathrm{E}_{\gamma_1,\gamma_2}[\sum_{t=0}^{\infty} \beta^t \sum_{k=1}^{\lambda_1} \sum_{\ell=1}^{\lambda_2} \mu_k^{\gamma_1(x(t))} \nu_\ell^{\gamma_2(x(t))} r(x(t), u_1^k, u_2^\ell) | x(0) = s_0].$$

Player 2's payoff is equal to $-V(s_0; \gamma_1, \gamma_2)$.

Definition 5.2.1 *A pair of strategies (γ_1^*, γ_2^*) is a saddle point if, for all strategies γ_1 and γ_2 of Players 1 and 2 respecively, for all $s \in S$*

$$V(s; \gamma_1, \gamma_2^*) \leq V(s; \gamma_1^*, \gamma_2^*) \leq V(s; \gamma_1^*, \gamma_2). \tag{5.1}$$

The number

$$v^*(s) = V(s; \gamma_1^*, \gamma_2^*)$$

is called the value of the game at state s.

Memory strategies: Since the game is of perfect recall, the information on which a player can base his/her decision at period t is the whole state-history

$$h(t) = \{x(0), x(1), \ldots, x(t)\}.$$

Notice that we don't assume that the players have a direct access to the actions used in the past by their opponent; they can only observe the *state history*. However, as in the case of a single-player Markov decision processes, it can be shown that optimality does not require more than the use of Markov strategies.

5.3 Shapley's-Denardo operator formalism

We introduce, in this section the powerful formalism of dynamic progaming operators that has been formally introduced in [Denardo, 1967] but was already implicit in Shapley's work. The solution of the dynamic programming equations for the stochastic game is obtained as a fixed point of an operator acting on the space of value functions.

5.3.1 Dynamic programming operators

In a seminal paper [Shapley. 1953] the existence of optimal stationary Markov strategies has been established via a fixed point argument, involving a *contracting operator*. We give here a brief account of the method, taking our inspiration from [Denardo, 1967] and [Filar & Vrieze, 1997]. Let $v(\cdot) = (V(s) : s \in S)$ be an arbitrary function with real values, defined over S. Since we assume that S is finite this is also a vector. Introduce, for any $s \in S$ the so-called *local reward functions*

$$h(v(\cdot), s, u_1, u_2) = r(s, , u_1, u_2) + \beta \sum_{s' \in S} p_{s,s'}(u_1, u_2)\, v(s'), \quad (u_1, u_2) \in U_1 \times U_2.$$

Define now, for each $s \in S$, the zero-sum matrix game with pure strategies U_1 and U_1 and with payoffs

$$H(v(\cdot), s) = [h(v(\cdot), s, u_1^k, u_2^\ell)] \begin{array}{l} k = 1, \ldots, \lambda_1 \\ \ell = 1, \ldots, \lambda_2 \end{array}.$$

Denote the value of each of these games by

$$T(v(\cdot), s) := \mathrm{val}[H(v(\cdot), s)]$$

and let $\mathbf{T}(v(\cdot)) = (T(v(\cdot), s) : s \in S)$. This defines a mapping $\mathbf{T} : \mathbf{R}^n \mapsto \mathbf{R}^n$. This mapping is also called the *Dynamic Programming operator* of the Markov game.

5.3.2 Existence of sequential saddle points

Lemma 5.3.1 *If A and B are two matrices of same dimensions, then*

$$|\text{val}[A] - \text{val}[B]| \leq \max_{k,\ell} |a_{k,\ell} - b_{k,\ell}|. \tag{5.2}$$

The proof of this lemma is left as exercise 6.1.

Lemma 5.3.2 *If $v(\cdot)$, γ_1 and γ_2 are such that, for all $s \in S$*

$$\begin{aligned} v(s) &\leq (\text{resp. } \geq \text{resp. } =) \quad h(v(\cdot), s, \gamma_1(s), \gamma_2 s(s)) \\ &= r(s, \gamma_1(s), \gamma_s(s)) + \beta \sum_{s' \in S} p_{s,s'}(\gamma_1(s), \gamma_2 s(s))\, v(s') \end{aligned} \tag{5.3}$$

then

$$v(s) \leq (\text{resp. } \geq \text{resp. } =) \quad V(s; \gamma_1, \gamma_2). \tag{5.4}$$

Proof: The proof is relatively straightforward and consists in iterating the inequality (5.3). **QED**

We can now establish the following result

Theorem 5.3.1 *the mapping $\mathbf{T}$ is contracting in the "max"-norm*

$$\|v(\cdot)\| = \max_{s \in S} |v(s)|$$

and admits the value function introduced in Definition 5.2.1 as its unique fixed-point

$$v^*(\cdot) = \mathbf{T}(v^*(\cdot)).$$

Furthermore the optimal (saddle-point) strategies are defined as the mixed strategies yielding this value, i.e.

$$h(v^*(\cdot), s, \gamma_1^*(s), \gamma_2^*(s)) = \text{val}[H(v^*(\cdot), s)], \quad s \in S. \tag{5.5}$$

Proof: We first establish the contraction property. We use lemma 5.3.1 and the transition probability properties to establish the following inequalities

$$\|\mathbf{T}(v(\cdot)) - \mathbf{T}(w(\cdot))\| \leq \max_{s \in S} \left\{ \max_{u_1 \in U_1;\, u_2 \in U_2} \left| r(s, u_1, u_2) + \beta \sum_{s' \in S} p_{s,s'}(u_1, u_2)\, v(s') \right. \right.$$

$$
\begin{aligned}
& \left. \left. -r(s,u_1,u_2) - \beta \sum_{s' \in S} p_{s,s'}(u_1,u_2)\, w(s') \right| \right\} \\
= & \max_{s \in S; u_1 \in U_1;\, u_2 \in U_2} \left| \beta \sum_{s' \in S} p_{s,s'}(u_1,u_2)\, (v(s') - w(s')) \right| \\
\leq & \max_{s \in S; u_1 \in U_1;\, u_2 \in U_2} \beta \sum_{s' \in S} p_{s,s'}(u_1,u_2)\ |v(s') - w(s')| \\
\leq & \max_{s \in S; u_1 \in U_1;\, u_2 \in U_2} \beta \sum_{s' \in S} p_{s,s'}(u_1,u_2)\ \|v(\cdot) - w(\cdot)\| \\
= & \beta \, \|v(\cdot) - w(\cdot)\| \, .
\end{aligned}
$$

Hence $\mathbf{T}$ is a contraction, since $0 \leq \beta < 1$. By the Banach contraction theorem this implies that there exists a unique fixed point $v^*(\cdot)$ to the operator $\mathbf{T}$.

We now show that there exist stationary Markov strategies γ_1^*, γ_2^* for wich the saddle point condition holds

$$V(s;\gamma_1,\gamma_2^*) \leq V(s;\gamma_1^*,\gamma_2^*) \leq V(s;\gamma_1^*,\gamma_2). \tag{5.6}$$

Let $\gamma_1^*(s), \gamma_2^*(s)$ be the saddle point strategies for the local matrix game with payoffs

$$h(v^*(\cdot), s, u_1, u_2)_{u_1 \in U_1, u_2 \in U_2} = H(v(\cdot), s). \tag{5.7}$$

Consider any strategy γ_2 for Player 2. Then, by definition

$$h(v^*(\cdot), s, \gamma_1^*(s), \gamma_2(s))_{u_1 \in U_1, u_2 \in U_2} \geq v^*(s) \quad \forall s \in S. \tag{5.8}$$

By Lemma 5.3.2 the inequality (5.8) implies that for all $s \in S$

$$V(s;\gamma_1^*,\gamma_2) \geq v^*(s). \tag{5.9}$$

Similarly we would obtain that for any strategy γ_1 and for all $s \in S$

$$V(s;\gamma_1,\gamma_2^*) \leq v^*(s), \tag{5.10}$$

and

$$V(s;\gamma_1^*,\gamma_2^*) = v^*(s), \tag{5.11}$$

This establishes the saddle point property. **QED**

In the single player case (MDP or Markov Decision Process) the contraction and monotonicity of the dynamic programming operator yield readily a converging numerical algorithm for the computation of the optimal strategies [Howard, 1960]. In the two-player zero-sum case, even if the existence theorem can also be translated into a converging algorithm, some difficulties arise (see e.g. [Filar & Tolwinski, 1991] for a recently proposed converging algorithm).

Chapter 6

Nonzero-sum Markov and Sequential games

In this chapter we consider the possible extension of the Shapley Markov game formalism to nonzero-sum games. We shall consider two steps in these extensions: (i) use the same finite state and finite action formalism as in Shapley's and introduce different reward functions for the different players; (ii) introduce a more general framework where the state and actions can be continuous variables.

6.1 Sequential Game with Discrete state and action sets

We consider a stochastic game played non-cooperatively by players that are not in purely antagonistic sutuation. Shapley's formalism extends without difficulty to a nonzero-sum setting.

6.1.1 Markov game dynamics

Introduce $S = \{1, 2, \ldots, n\}$, the set of possible states, $U_j(s) = \{1, 2, \ldots, \lambda_j(s)\}$, $j = 1, \ldots, m$, the finite *action sets* at state s of m players and the transition probabilities

$$p_{s,s'}(\mathbf{u}) = \mathrm{P}[x(t+1) = s' | x(t) = s,\ \mathbf{u}], \quad s, s' \in S, \quad \mathbf{u} \in U_1(s) \times \ldots \times U_m(s).$$

Define the transition reward of player j when the process is in state s and the players take the action pair $\mathbf{u}$ by

$$r_j(s, \mathbf{u}), \quad s \in S, \quad \mathbf{u} \in U_1(s) \times \ldots \times U_m(s).$$

Remark 6.1.1 *Notice that we permit the action sets of the different players to depend on the current state s of the game. Shapley's theory remains valid in that case too.*

6.1.2 Markov strategies

Markov strategies are defined as in the zero-sum case. We denote $\mu_k^{\gamma_j(x(t))}$ the probability given to action u_j^k by player j when he uses strategy γ_j and the current state is x.

6.1.3 Feedback-Nash equilibrium

Define Markov strategies as above, i.e. as mappings from S into the players *mixed actions*. Player j's payoff is thus defined as

$$V_j(s_0;\gamma_1,\ldots,\gamma_m)=\mathrm{E}_{\gamma_1,\ldots,\gamma_m}[\sum_{t=0}^{\infty}\beta^t\sum_{k=1}^{\lambda_1}\ldots\sum_{\ell=1}^{\lambda_m}\mu_k^{\gamma_1(x(t))}\cdots\mu_\ell^{\gamma_m(x(t))}r_j(x(t),u_1^k,\ldots,u_m^\ell)|x(0)=s_0].$$

Definition 6.1.1 *An m-tuple of Markov strategies $(\gamma_1^*,\ldots,\gamma_m^*)$ is a feedback-Nash equilibrium if for all $s\in S$*

$$V_j(s;\gamma_1^*,\ldots,\gamma_j,\ldots,\gamma_m^*)\le V_j(s;\gamma_1^*,\ldots,\gamma_j^*,\ldots,\gamma_m^*)$$

for all strategies γ_j of player j, $j\in M$. The number

$$v_j^*(s)=V_j(s;\gamma_1^*,\ldots,\gamma_j^*,\ldots,\gamma_m^*)$$

will be called the equilibrium value of the game at state s for Player j.

6.1.4 Sobel-Whitt operator formalism

The first author to extend Shapley's work to a nonzero-sum framework has been Sobel [**?**]. A more recent treatment of nonzero-sum sequential games can be found in [Whitt, 1980].

We introduce the so-called *local reward functions*

$$h_j(s,v_j(\cdot),\mathbf{u}) \;=\; r_j(s,\mathbf{u})+\beta\sum_{s'\in S}p_{ss'}(\mathbf{u})v_j(s'),\quad j\in M \tag{6.1}$$

where the functions $v_j(\cdot) : S \mapsto \mathbf{R}$ are given *reward-to-go* functionals (in this case they are vectors of dimension $n = \mathrm{card}(S)$) defined for each player j. The local reward (6.1) is the sum of the transition reward for player j and the discounted expected reward-to-go from the new state reached after the transition. For a given s and a given set of reward-to-go functionals $v_j(\cdot),\ j \in M$, the local rewards (6.1) define a matrix game over the pure strategy sets $U_j(s),\ j \in M$.

We now define, for any given Markov policy vector $\gamma = \{\gamma_j\}_{j\in M}$, an operator H_γ, acting on the space of reward-to-go functionals, (i.e. n-dimensional vectors) and defined as

$$(H_\gamma \mathbf{v}(\cdot))(s) \quad = \quad \{\mathrm{E}_{\gamma(s)}[h_j(s, v_j(\cdot), \tilde{\mathbf{u}}]\}_{j\in M}. \tag{6.2}$$

We also introduce the operator F_γ defined as

$$(F_\gamma \mathbf{v}(\cdot))(s) \quad = \quad \{\sup_{\gamma_j} \mathrm{E}_{\gamma^{(j)}(s)}[h_j(s, v_j(\cdot), \tilde{\mathbf{u}}]\}_{j\in M}, \tag{6.3}$$

where $\tilde{\mathbf{u}}$ is the random action vector and $\gamma^{(j)}$ is the Markov policy obtained when only Player j adjusts his/her policy, while the other players keep their γ-policies fixed. In other words Eq. 6.3 defines the optimal reply of each player j to the Markov strategies chosen by the other players.

6.1.5 Existence of Nash-equilibria

The dynamic programming formalism introduced above, leads to the following powerful results (see [Whitt, 1980] for a recent proof)

Theorem 6.1.1 *Consider the sequential game defined above, then*

1. *The expected payoff vector associated with a stationary Markov policy γ is given by the unique fixed point $\mathbf{v}_\gamma(\cdot)$ of the contracting operator H_γ.*
2. *The operator F_γ is also contracting and thus admits a unique fixed-point $\mathbf{f}^\gamma(\cdot)$.*
3. *The stationary Markov policy γ^* is an equilibrium strategy iff*

$$\mathbf{f}^{\gamma^*}(s) \quad = \quad \mathbf{v}_{\gamma^*}(s), \quad \forall s \in S. \tag{6.4}$$

4. *There exists an equilibrium defined by a stationary Markov policy.*

Remark 6.1.2 *In the nonzero-sum case the existence theorem is not based on a contraction property of the "equilibrium" dynamic programming operator. As a consequence, the existence result does not translate easily into a converging algorithm for the numerical solution of these games (see [**?**]).*

6.2 Sequential Games on Borel Spaces

The theory of non-cooperative markov games has been extended by several authors to the case where the state and the actions are in continuous sets. Since we are dealing with stochastic processes, the apparatus of measure theory is now essential.

6.2.1 Description of the game

An m-player *sequential game* is defined by the following objects:

$$(S, \Sigma), U_j, \Gamma_j(\cdot), r_j(\cdot,\cdot), Q(\cdot,\cdot).\beta,$$

where:

1. (S, Σ) is a measurable state space with a countably generated σ-algebra of Σ subsets of S.
2. U_j is a compact metric space of actions for player j.
3. $\Gamma_j(\cdot)$ is a lower measurable map from S into nonempty compact subsets of U_j. For each s, $\Gamma_j(s)$ represents the set of admissible actions for player j.
4. $r_j(\cdot,\cdot) : S \times \mathbf{U} \mapsto \mathbf{R}$ is a bounded measurable *transition reward function* for player j. This functions are assumed to be continuous on $\mathbf{U}$, for every $s \in S$.
5. $Q(\cdot,\cdot)$ is a product measurable transition probability from $S \times \mathbf{U}$ to S. It is assumed that $Q(\cdot,\cdot)$ satisfies some regularity conditions which are too technical to be given here. We refer the reader to [Nowak, 199?] for a more precise statement.
6. $\beta \in (0,1)$ is the *discount factor*.

A *stationary Markov strategy* for player j is a measurable map $\gamma_j(\cdot)$ from S into the set $P(U_j)$ of probability measure on U_j such that $\gamma_j(s) \in P(\Gamma_j(s))$ for every $s \in S$.

6.2.2 Dynamic programming formalism

The definition of *local reward functions* given in 6.1 in the discrete state case has to be adapted to the continuous state format, it becomes

$$h_j(s, v_j(\cdot), \mathbf{u}) \;=\; r_j(s,\mathbf{u}) + \beta \int_S v_j(t) Q(dt|s,\mathbf{u}), \quad j \in M \tag{6.5}$$

The operators H_γ and F_γ are defined as above. The existence of equilibria is difficult to establish for this general class of sequential games. In [Whitt, 1980], the existence of ε-equilibria is proved, using an approximation theory in dynamic programming. The existence of equilibria was obtained only for special cases [1]

6.3 Application to a Stochastic Duopoloy Model

6.3.1 A stochastic repeated duopoly

Consider the stochastic duopoly model defined by the following linear demand equation

$$x(t+1) = \alpha - \rho[u_1(t) + u_2(t)] + \varepsilon(t)$$

which determines the price $x(t+1)$ of a good at period $t+1$ given the total supply $u_1(t) + u_2(t)$ decided at the end of period t by Players 1 and 2. Assume a unit production cost equal to γ. The profits, at the end of period t by both players (firms) are then determined as

$$\pi_j(t) = (x(t+1) - \delta)u_j(t).$$

Assume that the two firms have the same discount rate β, then over an infinite time horizon, the payoff to Player j will be given by

$$V_j = \sum_{t=0}^{\infty} \beta^t \pi_j(t).$$

This game is repeated, therefore an obvious equilibrium solution consists to play repeatedly the (static) Cournot solution

$$u_j^c(t) = \frac{\alpha - \delta}{3\rho}, \quad j = 1, 2 \tag{6.6}$$

which generates the payoffs

$$V_j^c = \frac{(\alpha - \delta)^2}{9\rho(1-\beta)} \quad j = 1, 2. \tag{6.7}$$

A symmetric Pareto (nondominated) solution is given by the repeated actions

$$u_j^P(t) = \frac{\alpha - \delta}{4\rho}, \quad j = 1, 2$$

[1] see [Nowak, 1985], [Nowak, 1987], [Nowak, 1993], [Parthasarathy Shina, 1989].

and the associated payoffs

$$V_j^P = \frac{(\alpha - \delta)}{8\rho(1-\beta)} \quad j = 1, 2.$$

where $\delta = \alpha - \gamma$.

The Pareto outcome dominates the Cournot equilibrium but it does not represent an equilibrium. The question is the following:

> *is it possible to construct a pair of memory strategies which would define an equilibrium with an outcome dominating the repeated Cournot strategy outcome and which would be as close as possible to the Pareto nondominated solution?*

6.3.2 A class of trigger strategies based on a monitoring device

The random perturbations affecting the price mechanism do not permit a direct extension of the approach described in the deterministic context. Since it is assumed that the actions of players are not directly observable, there is a need to proceed to some filtering of the sequence of observed states in order to *monitor* the possible breaches of agreement.

In [Green & Porter, 1985] a dominating memory strategy equilibrium is constructed, based on a *one-step memory* scheme. We propose below another scheme, using a *multistep memory*, that yields an outcome which lies closer to the Pareto solution.

The basic idea consists to extend the state space by introducing a new state variable, denoted v which is used to monitor a *cooperative policy* that all players have agreed to play and which is defined as $\phi : v \mapsto u_j = \phi(v)$. The state equation governing the evolution of this state variable is designed as follows

$$v(t+1) = \max\{-K, v(t) + x^e(t+1) - x(t+1)\}, \tag{6.8}$$

where x^e is the expected outcome if both players use the cooperative policy, i.e.

$$x^e(t+1) = \alpha - 2\rho\phi(v(t)).$$

It should be clear that the new state variable v provides a cumulative measure of the positive discrepancies between the expected prices x^e and the realized ones $x(t)$. The parameter $-K$ defines a lower bound for v. This is introduced to prevent a compensation of positive discrepancies by negative ones. A positive discrepancy could be an indication of *oversupply*, i.e. an indication that at least one player is not respecting the agreement and is maybe trying to take advantage over the other player.

If these discrepancies accumulate too fast, the evidence of *cheating* is mounting and thus some retaliation should be expected. To model the retaliation process we introduce another state variable, denoted y, which is a *binary* variable, i.e. $y \in \{0, 1\}$. This new state variable will be an indicator of the prevailing mood of play. If $y = 1$ then the game is played cooperatively; if $y = 0$, then the the game is played in a noncooperative manner, interpreted as a *punitive* or *retaliatory* mood of play.

This state variable is assumed to evolve according to the following state equation

$$y(t+1) = \begin{cases} 1 & \text{if} \quad y(t) = 1 \quad \text{and} \qquad v(t+1) < \theta(v(t)) \\ 0 & \text{otherwise,} \end{cases} \tag{6.9}$$

where the positive valued function $\theta : v \mapsto \theta(v)$ is a *design parameter* of this monitoring scheme.

According to this state equation, the cooperative mood of play will be maintained provided the cumulative positive discrepancies do not increase too fast from one period to the next. Also, this state equation tells us that, once $y(t) = 0$, then $y(t') \equiv 0$ for all periods $t' > t$, i.e a punitive mood of play lasts forever. In the models discussed later on we shall relax this assumption of everlasting punishment.

When the *mood of play* is noncooperative, i.e. when $y = 0$, both players use as a *punishment* (or *retaliation*) the static Cournot solution forever. This generates the expected payoffs V_j^c, $j = 1, 2$ defined in Eq. (6.7). Since the two players are identical we shall not use the subscript j anymore.

When the mood of play is cooperative, i.e. when $y = 1$, both players use an agreed upon policy which determines their respective controls as a function of the state variable v. This agreement policy is defined by the function $\phi(v)$. The expected payoff is then a function $W(v)$ of this state variable v.

For this agreement to be stable, i.e. not to provide a temptation to cheat to any player, one imposes that it be an equilibrium. Note that the game is now a *sequential Markov game* with a continuous state space. The dynamic programming equation characterizing an equilibrium is given below

$$\begin{aligned} W(v) \quad = \quad & \max_u \{[\alpha - \delta - \rho(\phi(v) + u)]u \\ & + \beta \mathrm{P}[v' \geq \theta(v)] V^c \\ & + \beta \mathrm{P}[v' < \theta(v)] \mathrm{E}[W(v') | v' < \theta(v)]\}, \end{aligned} \tag{6.10}$$

where we have denoted

$$v' = \max\{-K, v + \rho(u - \phi(v)) - \varepsilon\}$$

the random value of the state variable v after the transition generated by the controls $(u, \phi(v))$.

In Eq. (6.10) we recognize the immediate reward $[\alpha-\delta-\rho(\phi(v)+u)]u$ of Player 1 when he plays u while the opponent sticks to $\phi(v)$. This is added to the conditional expected payoffs after the transition to either the *punishment mood of play,* corresponding to the values $y = 0$ or the *cooperative mood of play* corresponding to $y = 1$.

A solution of these DP equations can be found by solving an associated fixed point problem, as indicated in [Haurie & Tolwinski, 1990]. To summarize the approach we introduce the operator

$$\begin{aligned}(T_\phi W)(v,u) &= [\alpha-\delta-\rho(u+\phi(v))]\,u+\beta(\alpha-\delta)^2\frac{F(s-\theta(v))}{9\rho(1-\beta)}\\ &\quad +\beta W(-K)[1-F(s-K)]\\ &\quad +\beta\int_{-K}^{\theta(v)} W(\tau)f(s-\tau)\,d\tau \end{aligned} \tag{6.11}$$

where $F(\cdot)$ and $f(\cdot)$ are the cumulative distribution function and the density probability function respectively of the random disturbance ε. We have also used the following notation

$$s = v+\rho(u-\phi(v)).$$

An equilibrium solution is a pair of functions $(w(\cdot), \phi(\cdot))$ such that

$$W(v) = \max_u T_\phi(W)(v,u) \tag{6.12}$$

$$W(v) = (T_\phi W)(v,\phi(v)). \tag{6.13}$$

In [Haurie & Tolwinski, 1990] it is shown how an adapatation of the Howard *policy improvement algorithm* [Howard, 1960] permits the computation of the solution of this sort of fixed-point problem. The case treated in [Haurie & Tolwinski, 1990] corresponds to the use of a quadratic function $\theta(\cdot)$ and a Gaussian distribution law for ε. The numerical experiments reported in [Haurie & Tolwinski, 1990] show that one can define, using this approach, a subgame perfect equilibrium which dominates the repeated Cournot solution.

In [Porter, 1983], this problem has studied in the case where the (inverse) demand law is subject to a multiplicative noise. Then one obtains an existence proof for a dominating equilibrium based on a simple *one-step memory* scheme where the variable v satisfies the following equation

$$v(t+1) = \frac{x^e - x(t+1)}{x(t)}.$$

This is the case where one does not monitor the cooperative policy through the use of a cumulated discrepancy function but rather on the basis of repeated identical tests. Also in Porter's approach the punishment period is finite.

In [Haurie & Tolwinski, 1990] it is shown that the approach could be extended to a full fledged Markov game, i.e. a sequential game rather than a repeated game. A simple model of Fisheries management was used in that work to illustrate this type of sequential game *cooperative* equilibrium.

6.3.3 Interpretation as a communication device

In our approach, by extending the state space description (i.e. introducing the new variables v and y), we retained a Markov game formalism for an extended game and this has permitted us to use dynamic programming for the characterization of subgame perfect equilibria. This is of course reminiscent of the concept of *communication device* considered in [Forges 1986]for repeated games and discussed in Part 1. An easy extension of the approach described above would lead to random transitions between the two moods of play, with transition probabilities depending on the monitoring statistic v. Also a punishment of random duration is possible in this model. In the next section we illustrate these features when we propose a differential game model with *random moods of play*.

The monitoring scheme is a communication device which receives as input the observation of the state of the system and sends as an output a public signal which is suggesting to play according to two different moods of play.

Part III

Differential games

Chapter 7

Controlled dynamical systems

7.1 A capital accumulation process

A firm is producing a good with some equipment that we call its *physical capital*, or more simply its *capital*. At time t we denote by $x(t) \in \mathbf{R}^+$ the amount of capital acccumulated by the firm. In the parlance of dynamical systems $x(t)$ is the *state variable*. A capital accumulation path over a time interval $[t^0, t^f]$ is therefore defined as the graph of the function $x(\cdot) : [t^0, t^f] \mapsto x(t) \in \mathbf{R}^+$. In the parlance of dynamical systems this will be also called a *state trajectory*.

The firm accumulates capital through the investment process. Denote by $u(t) \in \mathbf{R}^+$ the investment realised at time t. This corresponds to the *control variable* for the dynamical systems. We also assume that the capital depreciates (wears out) at a fixed rate of μ.

Let us denote $\dot{x}(t)$ the derivative over time $\frac{dx(t)}{dt}$. The evolution over time of the accumulated capital is then described by a differential equation, with initial data x^0, called the *state equation* of the dynamical system

$$\dot{x}(t) = u(t) - \mu\, x(t) \tag{7.1}$$
$$x(t^0) = x^0. \tag{7.2}$$

Now, once the firm has chosen an investment schedule $u(\cdot) : [t^0, t^f] \mapsto u(t) \in \mathbf{R}^+$, and the initial capital stock $x^0 \in \mathbf{R}^+$ being given, there exists a unique capital accumulation path that is determined as the unique solution of the differential equation (7.1) with initial data (7.2). It is easy to check that since the investment rates are never negative, the capital stock will remain non-negative ($\in \mathbf{R}^+$).

7.2 State equations for controlled dynamical systems

In a more general setting a controlled dynamical system is described by the the state equation

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), t) && (7.3)\\ x(t^0) &= x^0, && (7.4)\\ u(t) &\in U \subset \mathbf{R}^m && (7.5) \end{aligned}$$

where $x \in \mathbf{R}^n$ is the state variable, $u \in \mathbf{R}^m$ is the control variable constrained to remain in the set $U \subset \mathbf{R}^m$, $f(\cdot,\cdot,\cdot) : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R} \to \mathbf{R}^n$ is the *velocity function* and $x^0 \in \mathbf{R}^n$ is the initial state. Under some regularity conditions that we shall describe shortly, if the controller chooses a control function $u(\cdot) : [t^0, t^f] \mapsto u(t) \in U$, it results a unique solution $x(\cdot) : [t^0, t^f] \mapsto u(t) \in \mathbf{R}^n$ to the differential equation (7.3), with initial condition (7.4). This solution $x(\cdot)$ is called the *state trajectory* generated by the control $u(\cdot)$ from the initial state x^0.

Indeed one must impose some regularity conditions for assuring that the fundamental theorem of existence and uniqueness of the solution to a system of differential equations can be invoked. We give below the simplest set of conditions. There are more general ones that will be needed later on, but for the moment we shall content ourselves with the following

7.2.1 Regularity conditions

1. The velovity function $f(x, u, t)$ is C^1 in x and continuous in u and t.

2. The control function $u(\cdot)$ is *piecewise continuous*.

7.2.2 The case of stationary systems

If the velocity function does not depend explicitly on t we say that the system is *stationary*. The capital accumulation process of section 7.1 is an example of a stationary system.

7.2.3 The case of linear systems

A linear stationary system is described by a linear state equation

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{7.6}$$
$$x(t^0) = x^0, \tag{7.7}$$

where $A : n \times n$ and $B : n \times m$ are given matrices.

For linear systems we can give the explicit form of the solution to Eqs. (7.6,7.7).

$$x(t) = \Phi(t, t^0)x^0 + \Phi(t, t^0)^{-1} \int_{t^0}^{t} \Phi(s, t^0)Bu(s)ds, \tag{7.8}$$

where $\Phi(t, t^0)$ is the $n \times n$ *transfer matrix* that verifies

$$\frac{d}{dt}\Phi(t, t^0) = A\Phi(t, t^0) \tag{7.9}$$
$$\Phi(t^0, t^0) = I_n. \tag{7.10}$$

Indeed, we can easily check in Eq. (7.8) that $x(t^0) = x^0$. Furthermore, if one differentiates Eq. (7.8) over time one obtains

$$\begin{aligned} \dot{x}(t) &= A\Phi(t, t^0)x^0 + A\Phi(t, t^0)^{-1} \int_{t^0}^{t} \Phi(s, t^0)Bu(s)ds \\ &\quad + \Phi(t, t^0)^{-1}\Phi(t, t^0)Bu(t) \qquad (7.11) \\ &= Ax(t) + Bu(t). \qquad (7.12) \end{aligned}$$

Therefore the function defined by (7.8) satisfies the differential equation (7.6) with initial condition (7.7).

7.3 Feedback control and the stability issue

An interesting way to implement a control on a dynamical system described by the general state equation (7.3,7.4) is to define the control variable $u(t)$ as a function $\sigma(t, x(t))$ of the current time t and state $x(t)$. We then say that the control is defined by the feedback law $\sigma(\cdot, \cdot) : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^m$, or more directly that one uses the feedback control σ. The associated trajectory will then be the solution to the differential equation

$$\dot{x}(t) = F_\sigma(t, x(t)) \tag{7.13}$$
$$x(t^0) = x^0, \tag{7.14}$$

where we use the notation

$$F_\sigma(t, x(t)) = f(x(t), \sigma(t, x(t)), t).$$

We see that, if $F_\sigma(\cdot, \cdot)$ is regular enough, the feedback control σ will determine uniquely an associated trajectory emanating from x^0.

7.3.1 Feedback control of stationary linear systems

Consider the linear system defined by the state equation (7.6,7.7) and suppose that one uses a linear feedback law

$$u(t) = K(t)\, x(t), \tag{7.15}$$

where the matrix $K(t) : m \times n$ is called the *feedback gain* at time t. The controlled system is now described by the dufferential equation

$$\begin{aligned} \dot{x}(t) &= (A + BK(t))\, x(t) && (7.16)\\ x(t^0) &= x^0. && (7.17) \end{aligned}$$

7.3.2 stabilizing a linear system with a feedback control

When a stationary system is observed on an infinite time horizon it may be natural to use a constant feedback gain K. Stability is an important qualitative property of dynamical systems, observed over an infinite time horizon, that is when $t^f \to \infty$.

Definition 7.3.1 *The dynamical system (7.16,7.17) is*

asymptotically stable if

$$\lim_{t\to\infty} \|x(t)\| = 0; \tag{7.18}$$

globally asymptotically stable if (7.18) holds true for any initial condition x^0.

7.4 Optimal control problems

7.5 A model of optimal capital accumulation

Consider the system described in section 7.1, concerning a firm which accumulates a productive capital. The state equation describing this controlled accumulation process is (7.1,7.2

$$\begin{aligned} \dot{x}(t) &= u(t) - \mu\, x(t)\\ x(t^0) &= x^0, \end{aligned}$$

where $x(t)$ and $u(t)$ are the capital stock and the investment rate at time t, respectively. Assume that the output $y(t)$, produced by the firm at time t is given by a *production*

function $y(t) = \phi(x(t))$. Assume now that this output, considered as an homogenous good, can be either consumed, that is distributed as a profit to the owner of the firm, at a rate $c(t)$, or invested, at a rate $u(t)$. Thus one must have at any instant t

$$c(t) + u(t) = y(t) = \phi(x(t)). \tag{7.19}$$

Notice that we permit negative values for $c(t)$. This would correspond to an operation of borrowing to invest more. The owner wants to maximise the total profit accumulated over the time horizon $[t^0, t^f]$ which is given by the integral payoff

$$\begin{aligned} J(x^0; x(\cdot), u(\cdot)) &= \int_{t^0}^{t^f} c(t)\, dt \\ &= \int_{t^0}^{t^f} [\phi(x(t)) - u(t)]\, dt. \end{aligned} \tag{7.20}$$

We have used the notation $J(x^0; x(\cdot), u(\cdot))$ to emphasize the fact that the payoff depends on the initial capital stock and the investment schedule, which will determine the capital accumulation path, that is the trajectory $x(\cdot)$. The search for the optimal investment schedule can now be formulated as follows

$$\max_{x(\cdot), u(\cdot)} \quad \int_{t^0}^{t^f} [\phi(x(t)) - u(t)]\, dt \tag{7.21}$$

s.t.

$$\dot{x}(t) = u(t) - \mu\, x(t) \tag{7.22}$$

$$x(t^0) = x^0 \tag{7.23}$$

$$u(t) \geq 0 \tag{7.24}$$

$$t \in [t^0, t^f].$$

Notice that, in this formulation we consider that the optimization is performed over the set of admissible control functions and state trajectories that must satisfy the constraints given by the state equation, the initial state and the non-negativity of the investment rates, at any time $t \in [t^0, t^f]$. The optimisation problem (7.21-7.24) is an instance of the generic optimal control problem that we describe now.

7.6 The optimal control paradigm

We consider the dynamical system introduced in section 7.2 with a criterion function

$$J(t^0, x^0; x(\cdot), u(\cdot)) = \int_{t^0}^{t^f} L(x(t), u(t), t)\, dt \tag{7.25}$$

that is to be maximised through the appropriate choice of an admissible control function. This problem can be formulated as follows

$$\max_{x(\cdot), u(\cdot)} \quad \int_{t^0}^{t^f} L(x(t), u(t), t)\, dt \tag{7.26}$$

$$
\begin{array}{rcll}
\text{s.t.} & & & \\
\dot{x}(t) & = & f(x(t), u(t), t) & (7.27) \\
x(t^0) & = & x^0 & (7.28) \\
u(t) & \in & U \subset \mathbf{R}^m & (7.29) \\
t & \in & [t^0, t^f]. &
\end{array}
$$

7.7 The Euler equations and the Maximum principle

We assume here that $U = \mathbf{R}^m$ and that the functions $f(x, u, t)$ and $L(x, u, t)$ are both C^1 in x and u. Under these more stringrnt conditions we can derive necessary conditions for an optimal control by using some simple arguments from the calculus of variations. These consitions are also called *Euler equations* as they are equivalent to the celebrated conditions of Euler for the classical variational problems.

Definition 7.7.1 *The functions $\delta x(\cdot)$ and $\delta u(\cdot)$ are called variations locally compatible with the constraints if they satisfy*

$$
\begin{array}{rcll}
\dot{\delta} x(t) & = & f_x(x(t), u(t), t)\, \delta x(t) + f_u(x(t), u(t), t)\, \delta u(t)) & (7.30) \\
\delta x(t^0) & = & 0. & (7.31)
\end{array}
$$

Lemma 7.7.1 *If $u^*(\cdot)$ and $x^*(\cdot)$ are the optimal control and the associated optimal trajectories, then the differential at $(x^*(\cdot), u^*(\cdot))$ of the performance criterion must be equal to 0 for all variations that are locally compatible with the constraints.*

The differential at $(x^*(\cdot), u^*(\cdot))$ of the integral performance criterion is given by

$$
\delta J(x^0; x^*(\cdot), u^*(\cdot)) = \int_{t^0}^{t^f} [L_x(x^*(t), u^*(t), t)\, \delta x(t) + L_u(x^*(t), u^*(t), t)\, \delta u(t)]\, dt. \tag{7.32}
$$

In order to express that we only consider variations $\delta x(\cdot)$ and $\delta u(\cdot)$ that are compatible with the constraints, we say that $\delta u(\cdot)$ is chosen arbitrarily and $\delta x(\cdot)$ is computed according to the "variational equations" (7.30,7.31).

Now we use a "trick". We introduce an auxiliary function $\lambda(\cdot) : [t^0, t^f] \to \mathbf{R}^n$, where $\lambda(t)$ is of the same dimension as the state variable $x(t)$ and will be called the *co-state variable*. If Eq. 7.30 is verified we may thus rewrite the expression (7.32) of δJ as follows

$$
\delta J = \int_{t^0}^{t^f} [L_x(x^*(t), u^*(t), t)\, \delta x(t) + L_u(x^*(t), u^*(t), t)\, \delta u(t) +
$$

$$\lambda(t)'\{-\dot{\delta}x(t) + f_x(x^*(t), u^*(t), t)\,\delta x(t) + f_u(x^*(t), u^*(t), t)\,\delta u(t))\}\,dt. \tag{7.33}$$

We integrate by parts the expression in $\lambda(t)'\,\dot{\delta}x(t)$ and obtain

$$\int_{t^0}^{t^f} \lambda(t)'\,\dot{\delta}x(t)\,dt = \lambda(t^f)'\,\delta x(t^f) - \lambda(t^0)'\,\delta x(t^0) - \int_{t^0}^{t^f} \dot{\lambda}(t)'\,\delta x(t)\,dt.$$

We can thus rewrite Eq. (7.33) as follows (we don't write anymore the arguments $(x^*(t), u^*(t), t)$)

$$\begin{aligned} \delta J &= -\lambda(t^f)'\,\delta x(t^f) + \lambda(t^0)'\,\delta x(t^0) \\ &\quad \int_{t^0}^{t^f} [\{L_x + \lambda(t)'\,f_x\delta x(t) + \dot{\lambda}(t)\}'\,\delta x(t) \\ &\quad +\{L_u + \lambda(t)'\,f_u\delta x(t)\}\,\delta u(t)]\,dt. \end{aligned} \tag{7.34}$$

Let us introduce the Hamiltonian function

$$H(\lambda, x, u, t) = L(x, u, t) + \lambda'\,f(x, u, t). \tag{7.35}$$

Then we can rewrite Eq. (7.34) as follows

$$\begin{aligned} \delta J &= -\lambda(t^f)'\,\delta x(t^f) + \lambda(t^0)'\,\delta x(t^0) \\ &\quad \int_{t^0}^{t^f} [\{H_x(\lambda(t), x^*(t), u^*(t), t) + \dot{\lambda}(t)\}'\,\delta x(t) \\ &\quad +H_u(\lambda(t), x^*(t), u^*(t), t)\,\delta u(t)]\,dt. \end{aligned} \tag{7.36}$$

As we can choose the function $\lambda(\cdot)$ as we like, we may impose that it satisfy the following *adjoint variational equations*

$$\dot{\lambda}(t)' = -H_x(\lambda(t), x^*(t), u^*(t), t) \tag{7.37}$$

$$\lambda(t^f) = 0. \tag{7.38}$$

With this particular choice of $\lambda(\cdot)$ the differential δJ writes simply as

$$\delta J = \lambda(t^0)'\,\delta x(t^0) + H_u(\lambda(t), x^*(t), u^*(t), t)\,\delta u(t)]\,dt. \tag{7.39}$$

Since $\delta x(t^0) = 0$ and $\delta u(t)$ is arbitrary, $\delta J = 0$ for all variations locally compatible with the constraints only if

$$H_u(\lambda(t), x^*(t), u^*(t), t) \equiv 0. \tag{7.40}$$

We can summarise Eqs. (7.39-7.43 in the following theorem

Theorem 7.7.1 *If $u^*(\cdot)$ and $x^*(\cdot)$ are the optimal control and the associated optimal trajectory, then there exists a function $\lambda(\cdot) : [t^0, t^f] \to \boldsymbol{R}^n$ which satisfies the adjoint variational equations*

$$\dot{\lambda}(t)' = -H_x(\lambda(t), x^*(t), u^*(t), t) \quad t \in [t^0, t^f] \tag{7.41}$$
$$\lambda(t^f) = 0, \tag{7.42}$$

and such that

$$H_u(\lambda(t), x^*(t), u^*(t), t) \equiv 0, \quad t \in [t^0, t^f], \tag{7.43}$$

where the hamiltonian function H is defined by

$$H(\lambda, x, u, t) = L(x, u, t) + \lambda' f(x, u, t), \quad t \in [t^0, t^f]. \tag{7.44}$$

These necessary optimality conditions, called the Euler equations, have been generalized by Pontryagin [**?**] for dynamical systems where the velocity

$$f(x, u, t)$$

and the payoff rate

$$L(x, u, t)$$

are C^1 un x, continuous in u and t, and where the control constraint set U is compact. The following theorem is the famous "Maximum principle" of the theory of optimal control.

Theorem 7.7.2 *If $u^*(\cdot)$ and $x^*(\cdot)$ are the optimal control and the associated optimal trajectory, then there exists a function $\lambda(\cdot) : [t^0, t^f] \to \boldsymbol{R}^n$ which satisfies the adjoint-variational equations*

$$\dot{\lambda}(t)' = -H_x(\lambda(t), x^*(t), u^*(t), t) \quad t \in [t^0, t^f] \tag{7.45}$$
$$\lambda(t^f) = 0, \tag{7.46}$$

and such that

$$H(\lambda(t), x^*(t), u^*(t), t) = \max_{u \in U} H(\lambda(t), x^*(t), u, t), \quad t \in [t^0, t^f], \tag{7.47}$$

where the hamiltonian function H is defined as in Eq. (7.44).

7.8 An economic interpretation of the Maximum Principle

It could be shown that the co-state variable $\lambda(t)$ indicates the sensitivity of the performance criterion to a marginal variation of the state $x(t)$. It can therefore be interpreted

as a marginal value of the state variable. Along the optimal trajectory, the Hamiltonian is therefore composed of two "values"; the current payoff rate $L(x^*(t), u^*(t), t)$ and the value of the current state modification $\lambda(t)' f(x^*(t), u^*(t), t)$. This is this global value that is maximised w.r.t. u at any point of the optimal trajectory.

The adjoint variational equations express the evolution of the marginal value of the state variables along the trajectory.

7.9 Synthesis of the optimal control

7.10 Dynamic programming and the optimal feedback control

Theorem 7.10.1 *Assume that there exist a function $V^*(\cdot,\cdot) : \boldsymbol{R} \times \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ and a feedback law $\sigma^*(t,x)$ that satisfy the following functional equations*

$$
\begin{aligned}
-\frac{\partial}{\partial t}V^*(t,x) &= \max_{u\in U} H(\frac{\partial}{\partial x}V^*(t,x), x^*(t), u, t), \quad \forall t, \forall x && (7.48)\\
&= H(\frac{\partial}{\partial x}V^*(t,x), x^*(t), \sigma^*(t,x), t), \quad \forall t, \forall x && (7.49)\\
V^*(t^f, x) &= 0. && (7.50)
\end{aligned}
$$

Then $V^*(t^i, x^i)$ is the optimal value of the performance criterion, for the optimal control problem defined with initial data $x(t^i) = x^i$, $t^i \in [t^0, t^f]$. Furthermore the solution of the maximisation in the R.H.S. of Eq. (7.48) defines the optimal feedback control.

Proof:

Remark 7.10.1 *Notice here that the partial derivatives $\frac{\partial}{\partial x}V^*(t,x)$ appearing in the Hamiltonian in the R.H.S. of Eq. (7.48) is consistent with the interpretation of the co-state variable $\lambda(t)$ as the sensitivity of the optimal payoff to marginal state variations.*

7.11 Competitive dynamical systems

7.12 Competition through capital accumulation

7.13 Open-loop differential games

$$J_j(t^0, x^0; x(\cdot), u_1(\cdot), \ldots, u_m(\cdot)) \;=\; \int_{t^0}^{t^f} L_j(x(t), u_1(t), \ldots, u_m(t), t)\, dt \tag{7.51}$$

s.t.

$$\dot{x}(t) = f(x(t), u_1(t), \ldots, u_m(t), t) \tag{7.52}$$

$$x(t^0) = x^0 \tag{7.53}$$

$$u_j(t) \in U_j \subset \mathbf{R}^{m_j} \tag{7.54}$$

$$t \in [t^0, t^f].$$

7.13.1 Open-loop information structure

Each player knows the initial data t^0, x^0 and all the other elements of the dynamical system; Player j selects a control function $u_j(\cdot) : [t^0, t^f] \to U_j$; the game is played as a single simultaneous move.

7.13.2 An equilibrium principle

Theorem 7.13.1 *If $\mathbf{u}^*(\cdot)$ and $x^*(\cdot)$ are the Nash equilibrium controls and the associated optimal trajectory, then there exists m functions $\lambda_j(\cdot) : [t^0, t^f] \to \boldsymbol{R}^n$ which satisfy the adjointvariational equations*

$$\dot{\lambda}_j(t)' = -H_{j\,x}(\lambda_j(t), x^*(t), \mathbf{u}^*, t) \quad t \in [t^0, t^f] \tag{7.55}$$

$$\lambda_j(t^f) = 0, \tag{7.56}$$

and such that

$$H_j(\lambda_j(t), x^*(t), \mathbf{u}^*, t) = \max_{u_j \in U_j} H(\lambda(t), x^*(t), [\mathbf{u}^{*-j}, u_j], t), \quad t \in [t^0, t^f], \tag{7.57}$$

where the hamiltonian function H_j is defined as follows

$$H_j(\lambda_j, x, \mathbf{u}, t) = L(x, \mathbf{u}, t) + \lambda_j' f(x, \mathbf{u}, t). \tag{7.58}$$

7.14 Feedback differential games

7.14.1 Feedback information structure

Player j can observe time and state $(t, x(t))$; he defines his control as the result of a feedback law $\sigma_j(t, x)$; the game is played as a single simultaneous move where each player announces the strategy he will use.

7.14.2 A verification theorem

Theorem 7.14.1 *Assume that there exist m functions $V_j^*(\cdot,\cdot) : \boldsymbol{R} \times \boldsymbol{R}^n \to \boldsymbol{R}$ and m feedback strategies $\sigma^{*-j}(t, x)$ that satisfy the following functional equations*

$$-\frac{\partial}{\partial t}V_j^*(t,x) = \max_{u_j \in U} H_j(\frac{\partial}{\partial x}V_j^*(t,x), x^*(t), [\sigma^{*-j}, u_j], t), \quad t \in [t^0, t^f], x \in \boldsymbol{R}^n \tag{7.59}$$

$$V_j^*(t^f, x) = 0. \tag{7.60}$$

Then $V_j^*(t^i, x^i)$ is the equilibrium value of the performance criterion of Player j, for the feedback Nash differential game defined with initial data $x(t^i) = x^i$, $t^i \in [t^0, t^f]$. Furthermore the solution of the maximisation in the R.H.S. of Eq. (7.59) defines the equilibrium feedback control of Player j.

7.15 Why are feedback Nash equilibria outcomes different from Open-loop Nash outcomes?

7.16 The subgame perfectness issue

7.17 Memory differential games

7.18 Characterizing all the possible equilibria

Part IV

A Differential Game Model

in this last part of our presentation we propose a stochastic differential games model of economic competition through technological innovation ($R\&D$ competition) and we show how it is connected to the theory of Markov and/or sequential games discussed in Part 2. This example deals with a stochastic game where the random disturbances are modelled as a *controlled jump process*.

In control theory, an interesting class of stochastic systems has been studied, where the random perturbations appear as *controlled jump processes*. In [**?**], [**?**] the relationship between these control models and *discrete event* dynamic programming models is developed. In [**?**] a first adaptation of this approach to the case of dynamic stochastic games has been proposed. Below we illustrate this class of games through a model of competition through investment in $R\&D$.

7.19 A Game of R&D Investment

7.19.1 Dynamics of $R\&D$ competition

We propose here a model of competition through $R\&D$ which extends earlier formulations which can be found in [**?**]. Consider m firms competing on a market with differentiated products. Each firm can invest in $R\&D$ for the purpose of making a technological advance which could provide it with a competitive advantage. The competitive advantage and product differentiation concepts will be discussed later. We focuss here on the description of the $R\&D$ dynamics. Let x_j be the level of accumulation of $R\&D$ capital by the firm j. The state equation describing the capital accumulation process is

$$\left.\begin{array}{rcl} \dot{x}_j(t) & = & u_j(t) - \mu_j x_j(t) \\ x_j(0) & = & x_j^0, \end{array}\right\} \tag{7.61}$$

where the control variable $u_j \in U_j$ gives the investment rate in $R\&D$ and μ_j is the depreciation rate of $R\&D$ capital.

We represent firm's j policy of $R\&D$ investment as a function

$$u_j(\cdot) : [0, \infty) \mapsto U_j.$$

Denote $\mathcal{U}_j$ the class of admissible functions $u_j(\cdot)$. With an initial condition $x_j(0) = x_j^0$ and a piecewise continuous control function $u_j(\cdot) \in \mathcal{U}_j$ is associated a unique evolution $x_j(\cdot)$ of the de capital stock, solution of equation (7.61).

The $R\&D$ capital brings innovation through a discrete event stochastic process. Let T be a *random stopping time* which defines the date at which the advance takes place.

Consider first the case with a single firm j. The elementary conditional probability is given by

$$\mathrm{P}_{u_j(\cdot)}[T \in (t; t+dt) | T > t] \quad = \quad \omega_j(x_j(t))dt + o(dt). \tag{7.62}$$

where $\lim_{dt \to 0} \frac{o(dt)}{dt} = 0$ uniformly in x_j. The function $\omega_j(x_j)$ represents the controlled intensity of the jump process which describes the innovation. The probability of having a technological advance occuring between times t and $t + dt$ is given by

$$\mathrm{P}_{u_j(\cdot)}[t < T < t+dt] \quad = \quad \omega_j(x_j(t)) e^{-\int_0^t \omega_j(x_j(s))ds} dt + o(dt) \tag{7.63}$$

Therefore the probability of having the occurence before time τ is given by

$$\mathrm{P}_{u_j(\cdot)}[T \leq \tau] \quad = \quad \int_0^\tau \omega_j(x_j(t)) e^{-\int_0^t \omega_j(x_j(s))ds} dt. \tag{7.64}$$

Equations (7.62-7.64) define the dynamics of innovation occurences for firm j.

Since there are m firms the combined jump rate will be

$$\omega(\mathrm{x}(t)) = \sum_{j=1}^{m} \omega_j(x_j(t)).$$

Given that a jump occurs at time τ, the conditionnal probability that the innovation come from firm j is given by

$$\frac{\omega_j(x_j(\tau))}{\omega(\mathrm{x}(\tau))}.$$

The impact of innovation on the competitive hedge of the firm is now discussed.

7.19.2 Product Differentiation

We model the market with differentiated products as in [**?**]. We assume that each firm j markets a product characterized by a quality index $q_j \in \mathbf{R}$ and a price $p_j \in \mathbf{R}$. Consider first a static situation where N consumers are buying this type of goods. It is assumed that the (indirect) utility of a consumer buying variant j is

$$\tilde{V}_j = y - p_j + \theta q_j + \varepsilon_j, \tag{7.65}$$

where θ is the valuation of quality and the ε_j are i.i.d., with standard deviation μ, according to the double exponential distribution (see [**?**] for details). Then the demand for variant j is given by

$$\tilde{D}_j = N P_j \tag{7.66}$$

where

$$P_j = \frac{\exp[(\theta q_j - p_j)/\mu]}{\sum_{i=1}^{m} \exp[(\theta q_i - p_i)/\mu]}, \quad j = 1, \ldots, m. \tag{7.67}$$

This corresponds to a *multinomial logit* demand model.

In a dynamic framework we assume a fixed number N of consumers with a demand rate per unit of time given by (7.65-7.67). But now both the quality index $q_j(\cdot)$ and the price $p_j(\cdot)$ are function of t. Actually they will be the instruments used by the firms to compete on this market.

7.19.3 Economics of innovation

The economics of innovation is based on the costs and the advantages associated with the $R\&D$ activity.

$R\&D$ operations and investment costs. Let $L_j(x_j, u_j)$ be a function which defines the cost rate of $R\&D$ operations and investment for firm j.

Quality improvement. When a firm j makes a technological advance, the quality index of the variant increases by some amount. We use a random variable $\Delta q_j(\tau))$ to represent this quality improvement. The probability distribution of this random variable, specified by the cumulative probability function $F_j(\cdot, x_j)$, is supposed to also depend on the current amount of know-how, indicated by the capital $x_j(\tau)$. The higher this stock of capital is, the higher will be the probability given to important quality gains.

Adjustment cost. Let $\Phi_j(x_j)$ be a monotonous et decreasing function of x_j which represents the adjustment cost for taking advantage of the advance. The higher the capital x_j at the time of the innovation the lower will be the adjustment cost.

Production costs Assume that the marginal production cost c_j for firm j is constant and does not vary at the time of an innovation.

Remark 7.19.1 *We implicitly assume that the firm which benefits from a technological innovation will use that advantage. We could extend the formalism to the case where a firm stores its technological advance and delays its implementation.*

7.20 Information structure

We assume that the firms observe the state of the game at each occurence of the discrete event which represents a technology advance. This system is an instance where two dynamics, a fast one and a slow one are combined, as shown in the definition of the state variables.

7.20.1 State variables

Fast dynamics. The levels of capital accumulation $x_j(t)$, $j = 1, \ldots, m$ define the *fast moving* state of this system. These levels change according to the differential state equations (7.61).

Slow dynamics. The quality levels $q_j(t)$, $j = 1, \ldots, m$ define a *slow moving* state. These levels change according to the random jump process defined by the jump rates (7.61) and the distribution functions $F_j(\cdot, x_j)$. If firm j makes a technological advance at time τ then the quality index changes abruptly

$$q_j(\tau) = q_j(\tau^-) + \Delta q_j(\tau).$$

7.20.2 Piecewise open-loop game.

Assume that at each random time τ of technological innovation all firms observe the state of the game. i.e the vectors

$$\mathbf{s}(\tau) = (\mathbf{x}(\tau), \mathbf{q}(\tau)) = (x_j(\tau), q_j(\tau))_{j=1,\ldots,m}.$$

Then each firm j selects a price schedule $p_j(\cdot) : [\tau, \infty) \mapsto \mathbf{R}$ and an $R\&D$ investment schedule $u_j(\cdot) : [\tau, \infty) \mapsto \mathbf{R}$ which are defined as open-loop controls. These controls will be used until the next discrete event (technological advance) occurs. We are thus in the context of *piecewise deterministic games* as introduced in [**?**].

7.20.3 A Sequential Game Reformulation

In [**?**] it is shown how piecewise deterministic games can be studied as particular instances of sequential games with Borel state and action spaces. The fundamental element of a sequential game is the so-called *local reward functionals* which permit the definition of the dynamic programming operators.

Let $v_j(\mathbf{s})$ be the expected reward-to-go for player j when the system is in initial state $\mathbf{s}$. The local reward functional describes the total expected payoff for firm j when all firms play according to the open-loop controls $p_j(\cdot) : [\tau, \infty) \mapsto \mathbf{R}$ and $u_j(\cdot) : [\tau, \infty) \mapsto \mathbf{R}$ until the next technological event occurs and the subsequent rewards are defined by the $v_j(\mathbf{s})$ function. We can write

$$\begin{aligned} h_j(\mathbf{s}, v_j(\cdot), \mathbf{u}(\cdot), \mathbf{p}(\cdot)) &= \mathrm{E}_{\mathbf{s},\mathbf{u}(\cdot)} \Big[\int_0^\tau e^{-\rho t} \{ \tilde{D}_j(t)(p_j(t) - c_j) \\ &\quad - L_j(x_j(t), u_j(t)) \}\, dt + e^{-\rho\tau} v_j(\mathbf{s}(\tau)) \Big], \\ &\quad j = 1, \ldots, m, \end{aligned} \tag{7.68}$$

where ρ is the discount rate, τ is the random time of the next technological event and $\mathbf{s}(\tau)$ is the new random state reached right after the occurence of the next technological event, i.e. at time τ. From the expression (7.68) we see immediately that:

1. We are in a sequential game formalism, with continuous state space and functional sets as action spaces (∞-horizon controls).

2. The price schedule $p_j(t)$, $t \geq \tau$ can be taken as a constant (until the next technological event) and is actually given by the solution of the static oligopoly game with differentiated products of qualities $p_j(t)$, $j = 1, \ldots, m$. The solution of this oligopoly game is discussed in [**?**], where it is shown to be uniquely defined.

3. The $R\&D$ investment schedule is obtained as a trade-off between the random (due to the random stopping time τ) transition cost given by $\int_0^\tau e^{-\rho t} L_j(x_j(t), u_j(t))\, dt$ and the expected reward-to-go after the next technological event.

We shall limit our discussion to this definition of the model and its reformulation as a sequential game in Borel spaces. The sequential game format does not lend itself easily to qualitative analysis via analytical solutions. A numerical analysis seems more appropriate to explore the consequences of these economic assumptions. The reader will find in [**?**] more information about the numerical approximation of solutions for this type of games.

Bibliography

[Alj & Haurie, 1983] A. ALJ AND A. HAURIE, *Dynamic Equilibria in Multigeneration Stochastic Games*, **IEEE Trans. Autom. Control**, Vol. AC-28, 1983, pp 193-203.

[Aumann, 1974] R. J. AUMANN *Subjectivity and correlation in randomized strategies*, **J. Economic Theory**, Vol. 1, pp. 67-96, 1974.

[Aumann, 1989] R. J.AUMANN **Lectures on Game Theory**, Westview Press, Boulder *etc.* , 1989.

[Başar & Olsder, 1982] T. BAŞAR AND G. K. OLSDER, **Dynamic Noncooperative Game Theory**, Academic Press, New York, 1982.

[Başar, 1989] T. BAŞAR, *Time consistency and robustness of equilibria in noncooperative dynamic games* in F. VAN DER PLOEG AND A. J. DE ZEEUW EDS., **Dynamic Policy Games in Economics** , North Holland, Amsterdam, 1989.

[Başar, 1989] T. BAŞAR , *A Dynamic Games Approach to Controller Design: Disturbance Rejection in Discrete Time* in **Proceedings of the** 28^{th} **Conference on Decision and Control**, IEEE, Tampa, Florida, 1989.

[Bertsekas, 1987] D. P. BERTSEKAS **Dynamic Programming: Deterministic and Stochastic Models**, Prentice Hall, Englewood Cliffs, New Jersey, 1987.

[Brooke et al., 1992] A. BROOKE, D. KENDRICK AND A. MEERAUS, **GAMS. A User's Guide, Release 2.25**, Scientific Press/Duxbury Press, 1992.

[Cournot, 1838] A. COURNOT, **Recherches sur les principes mathématiques de la théorie des richesses**, Librairie des sciences politiques et sociales, Paris, 1838.

[Denardo, 1967] E.V. DENARDO, *Contractions Mappings in the Theory Underlying Dynamic Programming,* **SIAM Rev.** Vol. 9, 1967, pp. 165-177.

[Ferris & Munson, 1994] M.C. FERRIS AND T.S. MUNSON, *Interfaces to PATH 3.0*, to appear in **Computational Optimization and Applications**.

[Ferris & Pang, 1997a] M.C. FERRIS AND J.-S. PANG, **Complementarity and Variational Problems: State of the Art**, SIAM, 1997.

[Ferris & Pang, 1997b] M.C. FERRIS AND J.-S. PANG, *Engineering and economic applications of complementarity problems*, **SIAM Review**, Vol. 39, 1997, pp. 669-713.

[Filar & Vrieze, 1997] J.A. FILAR AND VRIEZE K., **Competitive Markov Decision Processes**, Springer-Verlag New York, 1997.

[Filar & Tolwinski, 1991] J.A. FILAR AND B. TOLWINSKI, *On the Algorithm of Pollatschek and Avi-Itzhak* in T.E.S. Raghavan et al. (eds) **Stochastic Games and Related Topics**, 1991, Kluwer Academic Publishers.

[Forges 1986] F. FORGES, 1986, *An Approach to Communication Equilibria*, **Econometrica**, Vol. 54, pp. 1375-1385.

[Fourer et al., 1993] R. FOURER, D.M. GAY AND B.W. KERNIGHAN, **AMPL: A Modeling Languagefor Mathematical Programming**, Scientific Press/Duxbury Press, 1993.

[Fudenberg & Tirole, 1991] D. FUDENBERG AND J. TIROLE **Game Theory**, The MIT Press 1991, Cambridge, Massachusetts, London, England.

[Haurie & Tolwinski, 1990] A. HAURIE AND B. TOLWINSKI *Cooperative Equilibria in Discounted Stochastic Sequential Games*, **Journal of Optimization Theory and Applications**, Vol. 64, No. 3, March 1990.

[Friedman 1977] J.W. FRIEDMAN, **Oligopoly and the Theory of Games**, Amsterdam: North-Holland, 1977.

[Friedman 1986] J.W. FRIEDMAN, **Game Theory with Economic Applications**, Oxford: Oxford University Press, 1986.

[Green & Porter, 1985] GREEN E.J. AND R.H. PORTER *Noncooperative collusion under imperfect price information* , **Econometrica**, Vol. 52, 1985, 1984, pp. 87-100.

[Harsanyi, 1967-68] J. HARSANY, *Games with incomplete information played by Bayesian players*, **Management Science**, Vol. 14, 1967-68, pp. 159-182, 320-334, 486-502.

[Haurie & Tolwinski 1985a] A. HAURIE AND B. TOLWINSKI, 1984, *Acceptable Equilibria in Dynamic bargaining Games*, **Large Scale Systems**, Vol. 6, pp. 73-89.

[Haurie & Tolwinski, 1985b] A. HAURIE AND B. TOLWINSKI, *Definition and Properties of Cooperative Equilibria in a Two-Player Game of Infinite Duration*, **Journal of Optimization Theory and Applications**, Vol. 46i, 1985, No.4, pp. 525-534.

[Haurie & Tolwinski, 1990] A. HAURIE AND B. TOLWINSKI, *Cooperative Equilibria in Discounted Stochastic Sequential Games*, **Journal of Optimization Theory and Applications**, Vol. 64, 1990, No.3, pp. 511-535.

[Howard, 1960] R. HOWARD, **Dynamic programming and Markov Processes**, MIT Press, Cambridge Mass, 1960.

[Moulin & Vial 1978] H. MOULIN AND J.-P. VIAL, *strategically Zero-sum Games: The Class of Games whose Completely Mixed Equilibria Cannot be Improved Upon*, **International Journal of Game Theory**, Vol. 7, 1978, pp. 201-221.

[Kuhn, 1953] H.W. KUHN, *Extensive games and the problem of information*, in H.W. Kuhn and A.W. Tucker eds. **Contributions to the theory of games**, Vol. 2, Annals of Mathematical Studies No 28, Princeton University pres, Princeton new Jersey, 1953, pp. 193-216.

[Lemke & Howson 1964] C.E. LEMKE AND J.T. HOWSON, *Equilibrium points of bimatrix games*, **J. Soc. Indust. Appl. Math.**, **12**, 1964, pp. 413-423.

[Lemke 1965] C.E. LEMKE, *Bimatrix equilibrium points and mathematical programming*, **Management Sci.**, **11**, 1965, pp. 681-689.

[Luenberger, 1969] D. G. LUENBERGER, **Optimization by Vector Space Methods**, J. Wiley & Sons, New York, 1969.

[Luenberger, 1979] D. G. LUENBERGER, **Introduction to Dynamic Systems: Theory, Models & Applications**, J. Wiley & Sons, New York, 1979.

[Mangasarian and Stone, 1964] O.L. MANGASARIAN AND STONE H., *Two-Person Nonzerosum Games and Quadartic progamming*, **J. Math. Anal. Applic.**,/bf 9, 1964, pp. 348-355.

[Merz, 1976] A. W. MERZ, *The Game of Identical Cars* in: **Multicriteria Decision Making and Differential Games**, G. Leitmann ed., Plenum Press, New York and London, 1976.

[Murphy,Sherali & Soyster, 1982] F.H. MURPHY,H.D.SHERALI AND A.L. SOYSTER, *A mathematical programming approach for determining oligopolistic market equilibrium*, **Mathematical Programming**, Vol. 24, 1982, pp. 92-106.

[Nash, 1951] J.F. NASH, *Non-cooperative games*, **Annals of Mathematics**, **54**, 1951, pp. 286-295.

[Nowak, 1985] A.S. NOWAK, *Existence of equilibrium Stationary Strategies in Discounted Noncooperative Stochastic Games with Uncountable State Space*, **J. Optim. Theory Appl.**, Vol. 45, pp. 591-602, 1985.

[Nowak, 1987] A.S. NOWAK, *Nonrandomized Strategy Equilibria in Noncooperative Stochastic Games with Additive Transition and Reward Structures*, **J. Optim. Theory Appl.**, Vol. 52, 1987, pp. 429-441.

[Nowak, 1993] A.S. NOWAK, *Stationary Equilibria for Nonzero-Sum Average Payoff Ergodic Stochastic Games with General State Space*, **Annals of the International Society of Dynamic Games**, Birkhauser, 1993.

[Nowak, 199?] A.S. NOWAK AND T.E.S. RAGHAVAN, *Existence of Stationary Correlated Equilibria with Symmetric Information for Discounted Stochastic Games* **Mathematics of Operations Research**, to appear.

[Owen, 1982] *Owen G.* **Game Theory**, Academic Press, New York, London *etc.* , 1982.

[Parthasarathy Shina, 1989] PARTHASARATHY T. AND S. SHINA, *Existence of Stationary Equilibrium Strategies in Nonzero-Sum Discounted Stochastic Games with Uncountable State Space and State Independent Transitions*, **International Journal of Game Theory**, Vol. 18, 1989, pp.189-194.

[Petit, 1990] M. L. PETIT, **Control Theory and Dynamic Games in Economic Policy Analysis**, Cambridge University Press, Cambridge *etc.* , 1990.

[Porter, 1983] R.H. PORTER, *Optimal Cartel trigger Strategies* , **Journal of Economic Theory**, Vol. 29, 1983, pp.313-338.

[Radner, 1980] R. RADNER, *Collusive Behavior in Noncooperative ε-Equilibria of Oligopolies with Long but Finite Lives*, **Journal of Economic Theory**, Vol.22, 1980, pp.136-154.

[Radner, 1981] R. RADNER, *Monitoring Cooperative Agreement in a Repeated Principal-Agent Relationship*, **Econometrica**, Vol. 49, 1981, pp. 1127-1148.

[Radner, 1985] R. RADNER, *Repeated Principal-Agent Games with Discounting*, **Econometrica**, Vol. 53, 1985, pp. 1173-1198.

[Rosen, 1965] J.B. ROSEN, *Existence and Uniqueness for concave n-person games*, **Econometrica**, **33**, 1965, pp. 520-534.

[Rutherford, 1995] T.F RUTHERFORD, *Extensions of GAMS for complementarity problems arising in applied economic analysis*, **Journal of Economic Dynamics and Control**, Vol. 19, 1995, pp. 1299-1324.

[Selten, 1975] R. SELTEN, *Rexaminition of the Perfectness Concept for Equilibrium Points in Extensive Games*, **International Journal of Game Theory**, Vol. 4, 1975, pp. 25-55.

[Shapley. 1953] L. SHAPLEY, *Stochastic Games, 1953*, **Proceedings of the National Academy of Science** Vol. 39, 1953, pp. 1095-1100.

[Shubik, 1975a] M. SHUBIK, **Games for Society, Business and War**, Elsevier, New York *etc.* 1975.

[Shubik, 1975b] M. SHUBIK, **The Uses and Methods of Gaming**, Elsevier, New York *etc.* , 1975.

[Simaan & Cruz, 1976a] M. SIMAAN AND J. B. CRUZ, *On the Stackelberg Strategy in Nonzero-Sum Games* in: G. LEITMANN ed., **Multicriteria Decision Making and Differential Games**, Plenum Press, New York and London, 1976.

[Simaan & Cruz, 1976b] M. SIMAAN AND J. B. CRUZ, *Additional Aspects of the Stackelberg Strategy in Nonzero-Sum Games* in: G. LEITMANN ed., **Multicriteria Decision Making and Differential Games**, Plenum Press, New York and London, 1976.

[Sobel 1971] M.J. SOBEL, *Noncooperative Stochastic games*, **Annals of Mathematical Statistics**, Vol. 42, pp. 1930-1935, 1971.

[Tolwinski, 1988] B. TOLWINSKI, **Introduction to Sequential Games with Applications**, *Discussion Paper No. 46*, Victoria University of Wellington, Wellington, 1988.

[Tolwinski, Haurie & Leitmann, 1986] B. TOLWINSKI, A. HAURIE AND G. LEITMANN, *Cooperative Equilibria in Differential Games*, **Journal of Mathematical Analysis and Applications**, Vol. 119,1986, pp.182-202.

[Von Neumann 1928] J. VON NEUMANN, *Zur Theorie der Gesellshaftsspiele*, **Math. Annalen**, **100**, 1928, pp. 295-320.

[Von Neumann & Morgenstern 1944] J. VON NEUMANN AND O. MORGENSTERN, **Theory of Games and Economic Behavior**, Princeton: Princeton University Press, 1944.

[Whitt, 1980] W. WHITT, *Representation and Approximation of Noncooperative Sequential Games*, **SIAM J. Control**, Vol. 18, pp. 33-48, 1980.

[Whittle, 1982] P. WHITTLE, **Optimization Over Time: Dynamic Programming and Stochastic Control**, Vol. 1, J. Wiley, Chichester, 1982.