Foundations and Trends® in Systems and Control
Vol. 1, No. 4 (2014) 379–522

DOI: 10.1561/2600000003

Game Theory: Models, Numerical Methods and Applications

Dario Bauso
Dipartimento di Ingegneria Chimica, Gestionale,
Informatica, Meccanica,
Università di Palermo
dario.bauso@unipa.it

Contents

Abstract

Game theory is the theory of "strategic thinking". Developed for military purposes and defense, in the past it has also been used as an alternative and complementary approach to deal with robustness in the presence of worst-case uncertainties or disturbances in many areas such as economics, engineering, computer science, just to name a few. However, game theory is recently gaining ground in systems and control engineering, mostly in engineered systems involving humans, where there is a trend to use game theoretic tools to design protocols that will provide incentives for people to cooperate. For instance, scientists tend to use game theoretic tools to design optimal traffic flows, or predicting or avoiding blackouts in power networks or congestion in cyber-physical networked controlled systems.

Incentives to cooperate are also crucial in dynamic resource allocation, multi-agent systems and social models (including social and economic networks). This paper assembles the material of two graduate courses given at the Department of Engineering Science of the University of Oxford in June-July 2013 and at the Department of Electrical and Electronic Engineering of Imperial College, in October-December 2013. The paper covers the foundations of the theory of noncooperative and cooperative games, both static and dynamic. It also highlights new trends in cooperative differential games, learning, approachability (games with vector payoffs) and mean–field games (large number of homogeneous players). The course emphasizes theoretical foundations, mathematical tools, modeling, and equilibrium notions in different environments.

D. Bauso. *Game Theory: Models, Numerical Methods and Applications.*
Foundations and Trends® in Systems and Control, vol. 1, no. 4, pp. 379–522, 2014.
DOI: 10.1561/2600000003.

1

Introduction

This first chapter is introductory and streamlines the foundations of the theory together with seminal papers and applications. The chapter introduces different types of games, such as simultaneous and sequential games, and the corresponding representations. In addition, it makes a clear distinction between cooperative and noncooperative games. The introduction proceeds with the formalization of fundamental notions like pure and mixed strategy, Nash equilibrium and dominant strategy (strategic/normal representation and extensive/tree representation). In the second part of this chapter, we pinpoint seminal results on the existence of equilibria. The end of the chapter is devoted to the illustration of classical games such as the Cournot duopoly, as an example of infinite game, or other stylized games in strategic form known as the coordination game, the Hawk and Dove game or the Stag-Hunt game. We make use of the Cournot duopoly to briefly discuss the iterated dominance algorithm.

1.1 Historical note, definitions and applications

The foundations of game theory are in the book [von Neumann and Morgenstern, 1944] by the mathematician John Von Neumann and the economist Oskar Morgenstern,

```
Theory of games and economic behavior,
Princeton University Press, 1944.
```

The book builds on prior research by von Neumann published in German [von Neumann, 1928]: *Zur Theory der Gesellschaftsspiele, Mathematische Annalen, 1928.* Quoting from [Aumann, 1987], *Morgenstern was the first economist clearly and explicitly to recognize that economic agents must take the interactive nature of economics into account when making their decisions. He and von Neumann met at Princeton in the late Thirties, and started the collaboration that culminated in the Theory of Games.*

Forerunners of the theory are considered the french philosopher and mathematician Antoine Augustin Cournot, who first introduced the "duopoly model" in 1838, and the german economist Heinrich Freiherr von Stackelberg, who formulated the equilibrium concept named after him in 1934 [von Stackelberg, 1934].

Game theory intersects several disciplines, see e.g., Table 1.1, and conventionally involves multiple players each one endowed with its own payoff. Thus, game theory is different from optimization where one has one single player who optimizes its own payoff. Game theory also differs from multi-objective optimization, the latter characterized by one player and multiple payoffs. In the '60s another discipline was founded dealing with multiple decision makers with a common payoff, known as team theory [Ho, 1980, Marschak and Radner, 1972, Bauso and Pesenti, 2012].

The literature provides several formal definitions of game theory. For instance, Maschler, Solan, and Zamir say that game theory is a methodology using *mathematical tools to model and analyze situations involving several decision makers (DMs), called players* [Maschler et al., 2013]. According to Osborne and Rubinstein game theory is *a bag of analytical tools designed to help us understand the phenomena that we*

	1 payoff	n payoffs
1 player	Optimization	Multi-objective optimization
n players	Team theory	Game theory

Table 1.1: A scheme relating game theory to other disciplines.

observe when DMs interact, (DMs are rational and reason strategically) [Osborne and Rubinstein, 1994]. Here, (individual) rationality and strategic reasoning mean that every DM *is aware of his alternatives, forms expectations about any unknowns, has clear preferences, and chooses his action deliberately after some process of optimization* [Osborne and Rubinstein, 1994]. Tijs in his book defines game theory as *a mathematical theory dealing with models of conflict and cooperation* [Tijs, 2003].

Game theoretic models arise in numerous application domains including:

Board and field games [Bewersdorff, 2004]. Board games like chess or draughts or field games such as football or rugby may admit a mathematical description via game theory, where the players' actions are elements of a given set, called actions' set, and the probability of win is the payoff that every player seeks to maximize. In rugby, for instance, certain tactics are successful only if the opponent is playing a certain tactic and thus the tactic choice is assimilated to a play of the rock-paper-scissors game. Theoretical foundations are to be found in *algorithmic game theory*, a research area intersecting algorithm design, game theory and artificial intelligence [Noam et al., 2007].

Marketing and commercial operations [Osborne and Rubinstein, 1990, Gibbons, 1992]. Competitive firms operating on a same market must be able to predict the impact of a new product. This involves a strategic analysis of the current market demand and of the reactions of the potential competitors in consequence of the introduction of the new product.

Politics [Morrow, 1994]. Here game theory provides useful indices to measure the power of parties involved in a governing coalition. Voting methods can also be rigorously analyzed through game theory. Regarding social policy making, game theory offers guidelines to governmental agencies to predict and analyze the impact of specific social policy choices, such as pension rules, education or labor reforms.

Defense [**Hamilton and Mesic, 2004**]. Game theory has contributed the notion of "strategic thinking" consisting in putting ourselves in the place of the opponent before making a decision, which is a milestone in the field of defense. Military applications related to missile pursuing fighter airplanes are also usually addressed using game theoretic models.

Robotics and multi-agent systems [**Shoham and Leyton-Brown, 2009**]. Here game theory provides models for the movement of automated robot vehicles with distributed task assignment. Path planning for robotic manipulation in presence of moving obstacles is also a classical game theory application.

Networks [Di Mare and Latora, 2007, Saad et al., 2009]. Game theory can be used to analyze the spread of innovation, or the propagation of opinions in social networks. In communication networks game theory is frequently used to study band allocations, and in security problems.

1.2 Types of games and representations

There are different types of games and corresponding representations. In this section, after providing a formal description of a game in generic terms, we distinguish between *cooperative* and *noncooperative*, *simultaneous* and *sequential* games and introduce the *strategic* or *normal* representation for the former games and the *extensive* or *tree* representation for the latter games.

1.2.1 What is a game?

A (strategic form) game is a tuple $\langle N, (\mathcal{A}_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where

- $N = \{1, 2, \ldots, n\}$ is the set of players (maximizers),
- $\mathcal{A}_i$ is the set of actions of player i,

- $A := \{a| \, a = (a_i)_{i \in N}, \, a_i \in \mathcal{A}_i, \, \forall i \in N\}$ is the set of action profiles,
- $u_i : A \to \mathbb{R}$ is the payoff function of player i, i.e.,

$$(a_1, \ldots, a_n) \mapsto u_i(a_1, \ldots, a_n).$$

Note that the payoff u_i is a profit (to maximize) but can also be a cost (to minimize).

An equivalent way of writing the action profiles is

$$(a_j)_{j \in N} = (a_1, \ldots, a_n) = (a_i, a_{-i}),$$

where $a_{-i} = (a_j)_{j \in N, j \neq i}$ is the action profile of all players except i.

1.2.2 Noncooperative vs. cooperative

A first major distinction is between noncooperative and cooperative game theory. In noncooperative games i) every player seeks its best response based on the available information and in order to maximize its own payoff, ii) there are no binding agreements on optimal joint actions, iii) pre-play communication is possibly allowed.

In cooperative games (which in turn divide into games with transferable (TU) and nontransferable (NTU) utilities) i) the players seek optimal joint actions (NTU), or reasonable cost/reward sharing rules (TU) that make the coalitions stable, ii) pre-play communication is allowed, and iii) side payments are also allowed (TU).

Note that while noncooperative game theory dominates almost every textbook in game theory, and is by far more widespread than cooperative game theory, there is a large consensus on the idea that cooperative game theory has a broader range of applications. Only recently cooperative game theory has attracted the attention of scientists from disciplines other than economics, and has become a major design tool in engineered systems [Saad et al., 2009].

Example 1.1. (***Prisoners' dilemma***) This is one of the most common and simple strategic models developed by Merrill Flood and Melvin Dresher of the RAND Corporation in 1950. The prison-sentence interpretation, and thus the corresponding name is due to Albert W. Tucker. The story is the following one: two criminals are arrested under the suspicion of having committed a crime for which the maximal sentence is four years. Each one may choose whether to cooperate with (C) or defect (D) the other fellow. If both defect (D,D), the sentence is mitigated to three years (each one gets one year of freedom). If both cooperate (C,C), the suspects are released after one year due to lack of evidence (each one gets three years of freedom). If only one cooperates, (C,D) or (D,C), the one who defects is released immediately (four years of freedom), while the other is sentenced to the maximal punishment (zero years of freedom). The game is represented in bimatrix form as displayed in Fig. 7.1. In a purely noncooperative context, every player

Cooperative

Non Cooperative

	C	*D*
C	(3,3)	(0,4)
D	(4,0)	(1,1)

Figure 1.1: Prisoners' dilemma: cooperative vs. noncooperative solutions.

will choose to defect (D) considering that he has no guarantee on the other's choice and therefore the resulting solution is (D, D). Differently, in a cooperative scenario, where both players can collude and negotiate joint actions, it is likely that both will end up cooperating (C, C).

1.2.3 Simultaneous vs. sequential games

A second major distinction is between simultaneous games and sequential games. In simultaneous games i) decisions are made once and for all and at the same time by all players; ii) there is no state nor any concept of strategy; iii) these games admit a common *representation in normal form* (also called *strategic* or *bimatrix form*). The Prisoners' dilemma is a classical example of simultaneous games in strategic form, see Fig. 1.1. Here the rows and the columns are associated to the actions or decisions of the players and the entries of the bimatrix are the payoffs. The representation does not carry any inbuilt information structure.

On the other hand, sequential games are those where i) one has a specific order of events, ii) as a consequence, the game has a state variable that collects information on earlier decisions, iii) latter players may know perfectly or imperfectly the actual state (full or partial information), iv) decisions are made depending on the state from which the notion of *strategy*, namely a mapping from states to actions, v) such games are conveniently represented in *extensive* or *tree form*, see Fig. 1.2.

Here the nodes are the states and are labeled with the player who is to act; the branches are the actions; the payoffs are associated to leaf nodes and depend on the whole history of actions taken. The representation has an inbuilt information structure. Fig. 1.2 depicts a two-player extensive game where player 1 plays first (stage 1) and can select either left L or right R. Player 2 plays second (stage 2) and can in turn select left l or right r in both states 1 and 2, which yields four possible actions l_1, r_1, l_2, r_2.

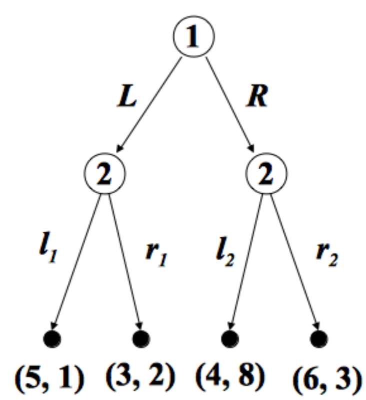

Figure 1.2: Example of extensive/tree form representation.

Nevertheless it is sometimes possible to derive an equivalent strategic form representation for an extensive game once we consider strategies rather than decisions. This is illustrated later on in Example 1.3 and Fig. 1.4. There we have four strategies for player 2, i.e., l_1l_2 (always left), l_1r_2 (left only in state 1, that is when player 1 picks L), l_2r_1 (left only in state 2, that is when player 1 picks R), and r_1r_2 (always right).

Thus, the set of "actions" for player 2 is $\mathcal{A}_2 = \{l_1l_2, l_1r_2, r_1l_2, r_1r_2\}$, while the one for player 1 is simply $\mathcal{A}_1 = \{L, R\}$.

Simultaneous games can also be played repeatedly over time in which case we address such games as *repeated games*. Repeated games admit an extensive form representation as shown below for the Prisoners' dilemma example in Fig. 1.3. Here payoffs or utilities are usually summed up over the rounds within a finite horizon or infinite horizon (discounted sum or long-term average) time window.

	C	**D**
C	(3,3)	(0,4)
D	(4,0)	(1,1)

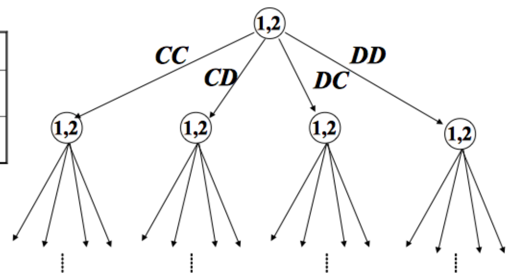

Figure 1.3: Extensive/tree form representation of the repeated Prisoners' dilemma.

1.3 Nash equilibrium and dominance

We review here basic solution concepts such as the Nash equilibrium and dominant strategy.

1.3.1 Nash equilibrium (NE)

In a Nash equilibrium "unilateral deviations" do not benefit any of the players. Unilateral deviations mean that only one player changes its own decision while the others stick to their current choices.

Definition 1.1. (Nash equilibrium [Nash Jr., 1950, 1951]) The action profile/outcome $(a_1^*, a_2^*, \ldots, a_n^*)$ is an NE if none of the players by deviating from it can gain anything, i.e.,

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \quad \forall a_i \in \mathcal{A}_i, \forall i \in N.$$

Let us introduce the best response set

$$\mathcal{B}_i(a_{-i}) := \{a_i^* \in \mathcal{A}_i | \, u_i(a_i^*, a_{-i}) = \max_{a_i \in \mathcal{A}_i} u_i(a_i, a_{-i})\}.$$

Then in an NE all players play a best response, namely

$$a_i^* \in \mathcal{B}_i(a_{-i}^*), \quad \forall i \in N.$$

Example 1.2. In the Prisoners' dilemma the solution (D, D) is a Nash equilibrium, as player 1 by deviating from it would get 0 years of freedom rather than 1 (stick to second column and move vertically to first row) and therefore would be worse off. Likewise for player 2.

	C	*D*
C	(3,3)	(0,4)
D	(4,0)	(1,1)

Example 1.3. In the extensive game of Fig. 1.4, player 2 has four strategies, i.e., $l_1 l_2$ (always left), $l_1 r_2$ (left only in state 1, that is when player 1 picks L), $l_2 r_1$ (left only in state 2, that is when player 1 picks R), and $r_1 r_2$ (always right). Thus, the set of "actions" for player 2 is $\mathcal{A}_2 = \{l_1 l_2, l_1 r_2, r_1 l_2, r_1 r_2\}$, while the one for player 1 is simply $\mathcal{A}_1 = \{L, R\}$. The game admits one Nash equilibrium $(R, r_1 l_2)$. This can be computed via dynamic programming backwardly. In state 1 (node down left), player 2's rational choice is r_1 (red line) as he gets 2 rather than 1 if he were to play l_1. In state 2, player 2 could play l_2 and get 8 or r_2 and get 3, then his rational choice is l_2 (red line). In stage 1 (top node), player one gets 4 by playing R, and 3 by playing L, so his best response is R (red line). The equilibrium payoffs are then $(4, 8)$.

- **E.g. Bimatrix Game**

	l_1l_2	l_1r_2	r_1l_2	r_1r_2
L	(5,1)	(5,1)	(3,2)	(3,2)
R	(4,8)	(6,3)	(4,8)*	(6,3)

Is this also a Nash equilibrium?

If Player 2 played l_1, L would be better than R for Player 1.

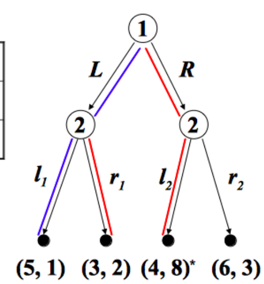

Figure 1.4: Nash equilibrium in an extensive tree game.

The representation in normal form of the game (left) shows another solution, $(R, l_1 l_2)$, returning the same payoffs as the equilibrium, which is not a Nash equilibrium as player 2 would benefit from changing from l_1 to r_1. So, in principle there may exist solutions that are not equilibria and which are equivalent to equilibria in terms of payoffs.

A weaker equilibrium solution concept is available in the literature, namely the so-called ϵ-Nash equilibrium.

Definition 1.2. **(ϵ-Nash equilibrium [Başar and Olsder, 1999, Chap. 4.2])** For a given $\epsilon \geq 0$, the action profile/outcome $(a_1^\epsilon, a_2^\epsilon, \ldots, a_n^\epsilon)$ is an ϵ-NE if none of the players by deviating from it can gain more than ϵ, i.e.,

$$u_i(a_i^\epsilon, a_{-i}^\epsilon) \geq u_i(a_i, a_{-i}^\epsilon) - \epsilon, \quad \forall a_i \in \mathcal{A}_i, \, \forall i \in N.$$

Needless to say, for $\epsilon = 0$, the ϵ-Nash equilibrium coincides with the Nash equilibrium.

1.3.2 Existence of equilibria and mixed strategies

The first seminal result on game theory is the minmax theorem by John Von Neumann, 1928, establishing the existence of equilibrium points for zero-sum games. These are games where the sum of the payoffs of the players is always zero. The result makes use of the notion of *mixed strategies*, namely strategies defined by a probability distribution over the finite set of the feasible strategies.

Theorem 1.1. **(Minmax theorem [von Neumann, 1928])** Each matrix game has a saddle point in the mixed strategies.

From a computational perspective, saddle points can be obtained via linear programming, which is the topic of Chap. 3 (see also Chap. 6, Tijs, 2003). The computation of NE is based on linear complementarity programming, which we will also discuss in Chap. 3 (see also Chap. 7, Tijs, 2003).

Existence of equilibria can be proven starting from the Kakutani's fixed point theorem 1941. The Kakutani's theorem analyzes sufficient conditions for a set-valued function, defined on a convex and compact subset of a Euclidean space, to have a *fixed point*, i.e. a point which is mapped to a set containing it.

Theorem 1.2. (**Kakutani's Fixed point theorem, 1941**) Let K be a non-empty subset of a finite dimensional Euclidean space. Let $f : K \to K$ be a correspondence, with $x \in K \mapsto f(x) \subseteq K$, satisfying the following conditions:

- K is a compact and convex set
- $f(x)$ is non-empty for all $x \in K$
- $f(x)$ is a convex-valued correspondence: for all $x \in K$, $f(x)$ is a convex set
- $f(x)$ has a closed graph: that is, if $\{x_n, y_n\} \to \{x, y\}$ with $y_n \in f(x_n)$, then $y \in f(x)$

Then, f has a fixed point, that is, there exists some $x \in K$, such that $x \in f(x)$.

Rather than the formal proof we provide a graphical illustration for a simple scalar case of the main ideas used in the proof. Let x be plotted in the horizontal axis, and $f(x)$ in the vertical axis as in Fig. 1.5. Fixed points, if exist, must solve $f(x) = x$ and therefore can be found at the intersection between the function $f(x)$ and the dotted line. On the left, the function $f(x)$ is not convex-valued and therefore it does not admit a fixed point. On the right, the function $f(x)$ does not have a closed graph which again implies that there exist no fixed point.

The Kakutani's theorem has been successively used by John Nash to prove the existence of a Nash equilibrium for nonzero-sum games. Essentially, the Nash's equilibrium theorem establishes the existence of at least one Nash equilibrium provided that i) the set of actions $\mathcal{A}_i$ are compact and convex subsets of $\mathbb{R}^n$, as it occurs in continuous (infinite) games, or games in mixed extension (we will expand more on it later); ii) payoffs $u_i(a_i, a_{-i})$ are continuous and concave in a_i for fixed strategy a_{-i} of the opponents.

Theorem 1.3. (Equilibrium point theorem [Nash Jr., 1950]) Each finite bimatrix game has an NE in the mixed strategies.

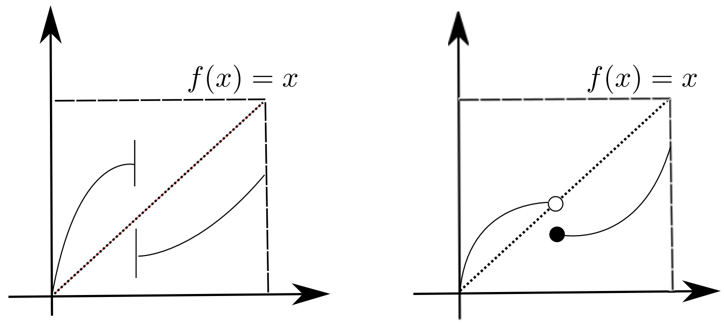

Figure 1.5: Graphical illustration of Kakutani's theorem. Function $f(x)$ is not convex valued (left), $f(x)$ has no closed graph (right). Courtesy by Asu Ozdaglar, slides of the course 6.254 Game Theory with Eng. Applications, MIT OpenCourseWare (2010).

Proof. We here provide only a sketch of the proof. Let us introduce the best response set,

$$\mathcal{B}_i(a_{-i}) := \{a_i^* \in \mathcal{A}_i |\, u_i(a_i^*, a_{-i}) = \max_{a_i \in \mathcal{A}_i} u_i(a_i, a_{-i})\}.$$

We can then apply the Kakutani's fixed point theorem to the best response correspondence $\mathcal{B} : \Delta \rightrightarrows \Delta$, $\Delta = \prod_{i \in N} \Delta_i$ (Δ_i is the simplex in the $\mathbb{R}^{|\mathcal{A}_i|}$)

$$\mathcal{B}(a) = \Big(\mathcal{B}_i(a_{-i})\Big)_{i \in N}.$$

□

An important property of mixed strategy Nash equilibria is that every action in the support of any player's equilibrium mixed strategy is a best response and yields that player the same payoff (cf. [Osborne and Rubinstein, 1994, Lemma 33.2]). We will henceforth refer to such a property as *Indifference Principle.*

Example 1.4. The example illustrated in Fig. 1.6 is borrowed from [Bressan, 2010] and describes a two-player continuous infinite game where the set of actions are segments in $\mathbb{R}$ (see horizontal and vertical axes). Level curves show that the maxima are attained at point P and Q for player 1 and 2 respectively. Note that the Nash equilibrium, which is point R, has horizontal and vertical tangents to the level curves

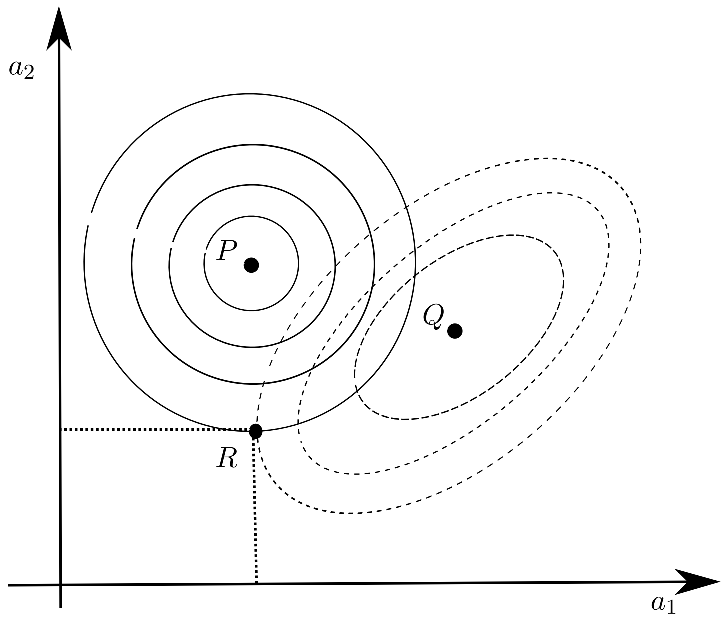

Figure 1.6: Two-player continuous infinite game. Level curves of player 1 (solid) and player 2 (dashed), action space of player 1 (horizontal axis), and of player 2 (vertical axis). Global maximum is P for player 1 and Q for player 2 while the NE is point R. Courtesy by Alberto Bressan, Noncooperative Differential Games. A Tutorial (2010) [Bressan, 2010].

of player 1 and 2 passing through it. From belonging to a horizontal tangent we know that the horizontal coordinate of point R is the best response of player 1 to player 2. Likewise, from belonging to a vertical tangent, the vertical coordinate of R is the best response of player 2 to player 1.

1.3.3 Dominant strategies

While the concept of equilibrium involves action profiles, the property of dominance is a characteristic related to a single action. Thus we say that an action profile is an NE, and that a given action is dominant. Dominance is a strong property, in that we know that an action profile made by dominant strategies is an NE but the converse is not true, i.e., we can have an NE that does not involve dominant strategies.

Definition 1.3. (Weak Dominance) Given two strategies, $a_i^*, a_i \in \mathcal{A}_i$, we say that a_i^* *weakly dominates* a_i if it is at least as good as a_i for all choices of the other players $a_{-i} \in \mathcal{A}_{-i}$,

$$u_i(a_i^*, a_{-i}) \geq u_i(a_i, a_{-i}), \quad \forall a_{-i} \in \mathcal{A}_{-i}.$$

If the above inequality holds strictly, then we say that a_i^* *(strictly) dominates* a_i.

Example 1.5. In the Prisoner's Dilemma, strategy D is a dominant strategy.

	C	*D*
C	(3,3)	(0,4)
D	(4,0)	(1,1)

Figure 1.7: D is a dominant strategy in the Prisoner's dilemma.

We say that a strategy is (weakly) dominant if it (weakly) dominates all other strategies. Note that a profile of dominant strategies is a Nash equilibrium. However, the converse is not true. Dominance is used in the renowned *iterated dominance algorithm*, which at each iteration eliminates subsets of dominated solutions by pruning the corresponding node in a so-called exploration tree. The term "pruning" is common in the parlance of combinatorial optimization, where the search for the optimum is graphically represented by an exploration tree. Here nodes describe families of solutions, and if, based on estimates, a family does not contain the optimum, then one says that the corresponding node is pruned. Back to the game theoretic setting, it is well known that a rationalizable/serially undominated strategy survives to the algorithm pruning. This is illustrated next for the Cournot duopoly.

1.4 Cournot duopoly and iterated dominance

Consider two manufacturers or firms $i = 1, 2$ competing on a same market. The production quantity of firm i is q_i. From the "law of demand", we assume that the sale price of firm i is $c_i = 30 - (q_1 + q_2)$.

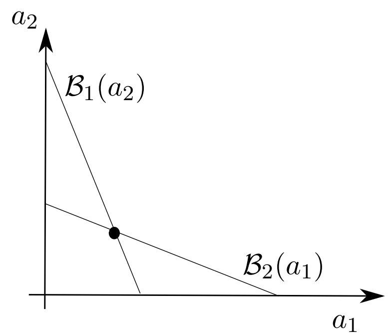

Figure 1.8: Reaction curves or best response for the Cournot duopoly.

Then, the payoff of firm i obtained by selling the produced quantity q_i at the price c_i is given by $u_i(q_i, q_j) = c_i q_i = 30q_i - q_i^2 - q_i q_j$. Note that this payoff is concave in q_i. Taking the derivative equal to zero, namely, $\frac{\partial u_i}{q_i} = 0$, it yields the best response $q_i^* = \mathcal{B}_i(q_j) = 15 - q_j/2$. The Nash equilibrium is $(10, 10)$, namely the intersection of the best response functions (see Fig. 1.8).

1.4.1 Iterated dominance algorithm

Figure 1.9 illustrates the iterated dominance algorithm on the Cournot duopoly model. To every iteration corresponds one round of elimination, which eliminates dominated strategies. Let us denote S_i^j the set of actions of player i that have survived to the elimination rounds up to iteration j. Then, one round of elimination yields $S_1^1 = [0, 1/2]$, $S_2^1 = [0, 1/2]$ (left). Indeed, player 1 knows that any production rate greater than $1/2$ is a dominated action for player 2 as $\mathcal{B}_2(a_1)$ lives in the range $[0, 1/2]$. We can repeat the same reasoning for player 2. Thus the search for equilibria can be restricted to the new domain $[0, 1/2]$ for both players (dotted square on the left plot). A second round yields $S_1^2 = [1/4, 1/2]$, $S_2^2 = [1/4, 1/2]$ (right). Actually, after restricting the best responses to the dotted square (left), every player knows that the best response of its opponent lives in the range $[1/4, 1/2]$. Thus the new search domain is $[1/4, 1/2]$ for both players which corresponds to the square (solid line) on the right plot. Repeating the same reasoning iteratively, the algorithm is proven to converge to the Nash equilibrium.

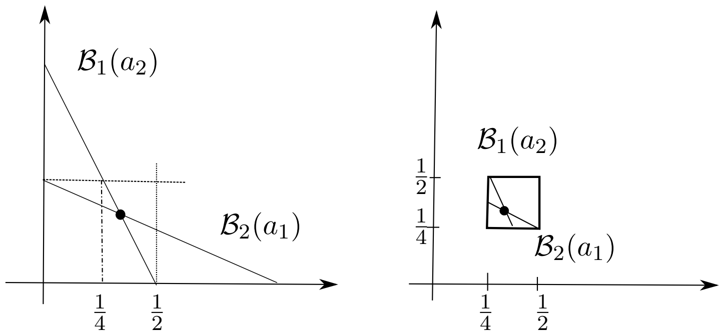

Figure 1.9: Iterated dominance algorithm illustrated on the Cournot duopoly model. Courtesy by Asu Ozdaglar, slides of the course 6.254 Game Theory with Eng. Applications, MIT OpenCourseWare (2010).

1.5 Examples

The rest of this chapter illustrates classical examples of strategic games such as: the *battle of the sexes*, the *coordination or typewriter game*, the *Hawk and dove or chicken game*, and the *Stag-Hunt game*.

Example 1.6. (Battle of the sexes) A couple agrees to meet in the evening either to go shopping S or to attend a cricket match C. The husband (column player) prefers to go to the cricket game while the wife (row player) would like to go shopping. In any case, both wish to go to the same place. Payoffs measure the happiness of the two. If they both go shopping, i.e. (SS), the woman is happy 2 and the husband is happy 1, while if they both go to the cricket game the happiness levels swap, 2 for the husband and 1 for the woman. If they end up in different places the level of happiness of both is 0.

	S	C
S	(2,1)	(0,0)
C	(0,0)	(1,2)

Example 1.7. (Coordination game or Typewriter game) Usually presented as a stylized model for diffusion of innovation, when and where is it convenient to adopt a new technology, this game considers a couple which agrees to meet in the evening either to go to a Mozart's or to a Mahler's concert. Both players have a small preference for Mozart

and if they both select $(Mozart, Mozart)$, then each level of happiness is 2. The levels of happiness are a bit lower, say 1, if they both go to a Mahler concert, i.e., $(Mahler, Mahler)$. Going to two different concerts returns a level of happiness equal to 0 to both. The action profiles $(Mozart, Mozart)$ and $(Mahler, Mahler)$ are both NE solutions.

	$Mozart$	$Mahler$
$Mozart$	(2,2)	(0,0)
$Mahler$	(0,0)	(1,1)

Example 1.8. (Hawk and dove or chicken game) The underlying idea is that while each player prefers not to give in to the other, the worst possible outcome occurs when both players do not yield. The game simulates a situation where two drivers drive towards each other and the one who swerves at the last moment is addressed “chicken”. The same game under the name “Hawk and Dove" describes a scenario where two contestants can choose a nonaggressive or aggressive attitude. The game was useful to illustrate the strategic scenario during the cold war and in particular in occasion of the Cuban Missile Crisis. The game is mostly meaningful in the case where the cost of fighting exceeds the prize of victory, i.e., $C > V > 0$. If both player decide for a nonaggressive behavior and share the prey, i.e. they opt for $(Dove, Dove)$, their mutual reward is half of the prize of victory, $V/2$. If one yields, $(Hawk, Dove)$ or $(Dove, Hawk)$ the winner gets the entire prey, V and the loser is left with zero reward. If both players are aggressive and end up fighting, $(Hawk, Hawk)$, each will pay a cost equal to half of the prize of victory subtracted to the cost of fight. The game admits two NE solutions, $(Hawk, Dove)$ and $(Dove, Hawk)$.

	$Hawk$	$Dove$
$Hawk$	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
$Dove$	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

Example 1.9. (Stag-Hunt game) Used to analyze and predict social cooperation, this game illustrates situations where two individuals can go out on a hunt and collaborate or not collaborate. Each hunter can decide to hunt a stag or hunt a hare without knowing what the other

is going to do. If both cooperate and go for a stag, $(Stag, Stag)$, they will share a large prey and each revenue is $3/2$. If both go for a hare $(Hare, Hare)$ the revenue to share is lower and equal to 1. If they go for different preys, $(Stag, Hare)$ or $(Hare, Stag)$ the one who goes for the smaller prey (the hare) gets the entire prey for himself, while the other is left with nothing as hunting a stag alone is not possible. He must have the cooperation of his partner in order to succeed. An individual can get a hare by himself, but a hare is worth less than a stag.

	$Stag$	$Hare$
$Stag$	$\left(\frac{3}{2}, \frac{3}{2}\right)$	(0,1)
$Hare$	(1,0)	$\left(1, 1\right)$

2

Zero-sum games

Two-person zero-sum games (TPZSGs) are the purest form of non-cooperative games. The spirit of such games is captured by the latin expression:

“Mors tua vita mea”.

Essentially, if one player wins 1 dollar the other looses 1 dollar and vice versa. TPZSGs are interesting as it is easy to check whether NE solutions, now called *saddle-points*, exist or not. In addition to this, though the saddle-point may not be unique, for such games all saddle-points enjoy the i) interchangeability property, and the ii) equal payoff property.

TPZSGs play a major role from a historical standpoint. Indeed, the first seminal result in game theory has been actually derived for TPZSGs and is known as the *min-max theorem* by J. Von Neumann (1928) [von Neumann, 1928, von Neumann and Morgenstern, 1944]. The result is obtained considering what is called the *mixed extension* of TPZSGs. By that we mean games where players play mixed strategies, namely, they choose probabilities over their action spaces and randomize their actions based on such probabilities. The well known *equilibrium point theorem* by J. Nash, 1950 represents the extension to nonzero-sum games of the min-max theorem [Nash Jr., 1950].

2.1 Two-person zero-sum games

In TPZSGs the sum of the payoffs is zero whatever the action profile is, namely

$$u_1 = -u_2.$$

As such they can be represented as matrix games. In other words, rather than a bimatrix we can simply use a matrix with a scalar entry representing the payoff u_2. Thus we henceforth call A the matrix of the game and let a_{ij} be its ij-th entry, representing the payoff resulting from player 1 playing the ith row and player 2 playing the jth column. Given matrix A, we then assume that the row player P_1 is the minimizer (maximizing u_1 is equivalent to minimizing u_2) and the column player P_2 is the maximizer. This is illustrated in Fig. 2.1 which shows a TPZSG represented by an $(m \times n)$-matrix, namely, player 1 has m actions while player 2 has n actions.

	P_2 (maximizer)			
P_1 (minimizer)	$(\cancel{1}, -1)$	$(\cancel{3}, -3)$	$\dots$	a_{1n}
	$(\cancel{5}, -5)$	a_{22}	$\dots$	a_{2n}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	a_{m1}	a_{m2}	$\dots$	a_{mn}

Figure 2.1: Matrix game representation of a TPZSG

2.2 On the existence of saddle-points

In TPZSGs NE solutions are saddle-points. Actually, any best opponent's response yields the worst own payoff. Saddle-points are related to *conservative strategies* (i^*, j^*), which are given by:

$$\begin{cases} \overline{J}(A) := \min_i \max_j a_{ij} & \text{loss ceiling} \\ \underline{J}(A) := \max_j \min_i a_{ij} & \text{gain floor} \end{cases}. \tag{2.1}$$

Essentially every player optimizes the worst-case scenario. The loss ceiling (upper bound) $\overline{J}(A)$ is obtained by maximizing over the columns thus obtaining the worst-payoffs for every choice of player 1, and then taking the minimum over the row (best worst-payoff). Similarly, the gain floor $\underline{J}(A)$ (lower bound) is obtained by minimizing over the rows and then taking the maximum over the columns.

In pure strategies (players' actions are discrete) there exists a main result which provides existence conditions for saddle-points.

Theorem 2.1. A saddle-point exists if the gain floor is equal to the loss ceiling, i.e.,

$$\underline{J}(A) = a_{i^* j^*} = \overline{J}(A).$$

Furthermore, if a saddle-point exists, this corresponds to both players playing conservative, and the equilibrium payoff is then $a_{i^* j^*}$.

Example 2.1. Figure 2.2 depicts a matrix game for which no saddle-point exists. The loss ceiling $\overline{J}(A) = 2$ and the gain floor $\underline{J}(A) = -1$. To see this, consider all player 1's actions and the corresponding best response for player 2. In the example, we call s_1 the strategy of player 1 and s_2 the strategy of player 2. The strategy or action sets are $A_i = \{-1, 0, 1\}$ for $i = 1, 2$. In particular, if player 1 plays $s_1 = -1$ (top or 1st row), then player 2 responds with $s_2 = 1$ (right or 3rd column) which yields the payoff $a_{13} = 4$ (red circle). Differently, if player 1 plays $s_1 = 0$ (middle or 2nd row), then player 2 responds with $s_2 = -1$ or $s_2 = 1$ (he can play equivalently left or right, namely 1st or 3rd column) which yields the payoff $a_{21} = a_{23} = 2$ (red circles on second row). Finally, if player 1 plays $s_1 = 1$ (bottom or 3rd row), then player 2 responds with $s_2 = -1$ (left or 1st column) which yields the payoff $a_{31} = 4$ (red circles on third row). Player 1 can then choose what to play based on the predictable reaction of player 2, and after comparing the three different scenarios he will select $s_1 = 0$ (middle or 2nd row) which returns the minimum payoff (both a_{21} or a_{23} are less than a_{13} and a_{31}). Thus the conservative strategy for player 1 is $i^* = s_1 = 0$ and the gain floor $\overline{J}(A) = 2$. Note that this represents an upper bound. To see this, note that if player 1 plays conservative then the payoff cannot exceed such a value. Also, note that the entire procedure corresponds to the min-max calculation expressed in the first line of (2.1). Repeating the same reasoning for player 2, we obtain that the conservative strategy for him is $j^* = 0$ (middle or 2nd column) and the gain floor is $\underline{J}(A) = -1$. Note that this is a lower bound. Actually, if player 2 plays conservative, then the payoff can never be lower than such a value whatever player 1 picks. Therefore the payoff corresponding to both players playing conservative is $a_{i^*j^*} = a_{22} = 0$. Note that this is not a saddle-point. Indeed, as loss ceiling and gain floor are different, from Theorem 2.1 we conclude that no saddle-point exists for this game.

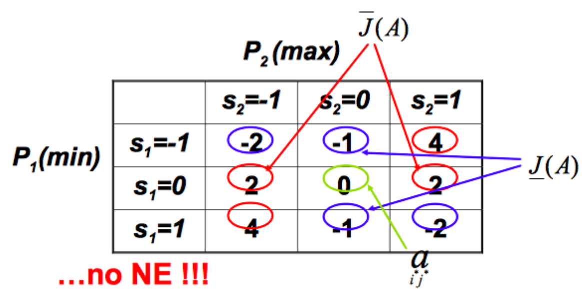

Figure 2.2: Loss ceiling $\overline{J}(A)$ and gain floor $\underline{J}(A)$; no saddle-point for this game.

Example 2.2. In this second example, the matrix game, which is illustrated in Fig. 2.3, admits a saddle-point. To see this, consider that whatever player 1 chooses, player 2 will always react by choosing the middle or 2nd column, see the red circles on each row. Then, the conservative strategy for player 1 is $i^* = s_1 = 0$ and the gain floor $\overline{J}(A) = 10$. Likewise, if player 2 plays $s_2 = -1$ (left or 1st column), then player 1 responds with $s_1 = -1$ (top or 1st row) which yields the payoff $a_{11} = -40$ (blue circle on first row). Differently, if player 2 plays $s_2 = 0$ (middle or 2nd column), then player 1 responds with $s_1 = 0$ (middle or 2nd column) which yields the payoff $a_{22} = 10$ (blue circles on second row). Finally, if player 2 plays $s_2 = 1$ (right or 3rd column), then player 1 responds with $s_1 = 1$ (bottom or 3rd row) which yields the payoff $a_{33} = -2$ (blue circle on third row). Player 2 can then choose what to play based on the predictable reaction of player 1, and after comparing the three different scenarios he will select $s_2 = 0$ (middle or

2nd row) which returns the maximum payoff (both a_{11} or a_{33} are less than a_{22}). Thus the conservative strategy for player 2 is $j^* = s_2 = 0$ and the gain floor $\overline{J}(A) = 10$. Therefore the payoff corresponding to both players playing conservative is $a_{i^*j^*} = a_{22} = 0$. As loss ceiling and gain floor coincide, one can see that Theorem 2.1 condition is satisfied and therefore $a_{i^*j^*}$ is also an equilibrium payoff.

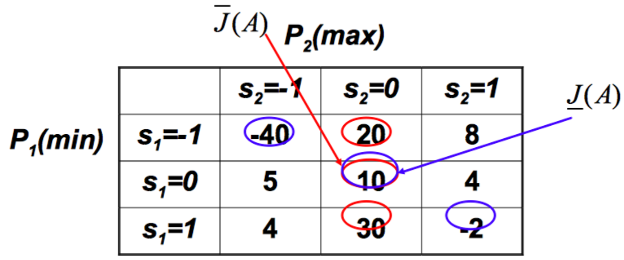

Figure 2.3: Loss ceiling $\overline{J}(A)$ and gain floor $\underline{J}(A)$; This game admits a saddle-point.

In general, there may exist multiple saddle-points. However, given two saddle points (i, j) and (k, l) we have:

- (interchangebility) (i, l) and (k, j) are also saddle-points
- (equal payoff) $a_{ij} = a_{kl} = a_{il} = a_{kj}$

Theorem 2.2. (Min-max [von Neumann, 1928, von Neumann and Morgenstern, 1944]): Each matrix game has a saddle point in the mixed strategies.

Saddle-points can be computed via linear programming as described in the next chapter.

2.3 Application: H^∞-optimal control

This section builds upon the introductory chapter of [Başar and Bernhard, 1995] and provides connections between zero-sum games and H^∞-optimal control. The latter is the theory supporting the design of controllers in the presence of worst-case uncertainty. Consider the block diagram in Fig. 2.4 depicting a plant G and a feedback controller

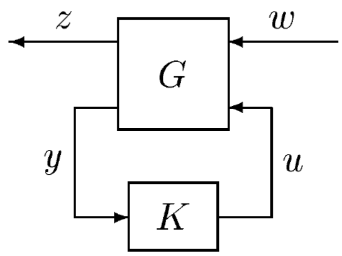

Figure 2.4: Block diagram of plant and feedback controller.

K. We also have a control u, a disturbance w, controlled and measured outputs z and y (all measurable in Hilbert spaces $\mathcal{H}_u, \mathcal{H}_w, \mathcal{H}_z, \mathcal{H}_y$). A classical representation of the dynamics of the system is given by

$$\begin{cases} z = G_{11}(w) + G_{12}(u), \\ y = G_{21}(w) + G_{22}(u), \\ u = K(y). \end{cases} \tag{2.2}$$

Here we assume that both the operators G_{ij} and the controller $K \in \mathcal{K}$ are bounded causal linear operators, where we denote by $\mathcal{K}$ the controller space. "Causal" means that all subsystems are nonanticipative, namely, the output may depend on past and current inputs but not on future inputs.

A main issue in robust control is related to the capability of the controlled plant to attenuate the effects of the disturbance, which is called *disturbance attenuation*. Such a problem can be converted into a zero-sum game between the controller and the disturbance. To see this, for every fixed $K \in \mathcal{K}$, introduce bounded causal linear operators $T_K : \mathcal{H}_w \to \mathcal{H}_z$

$$T_K(w) = G_{11}(w) + G_{12}(I - KG_{22})^{-1}(KG_{21})(w).$$

We then look for the worst-case infimum of the operator norm

$$\begin{cases} \inf_{K \in \mathcal{K}} \langle\langle T_K \rangle\rangle =: \gamma^*, \\ \langle\langle T_K \rangle\rangle = \sup_{w \in \mathcal{H}_w} \frac{\|T_K(w)\|_z}{\|w\|_w}. \end{cases} \tag{2.3}$$

This turns the problem into a TPZSG between controller and disturbance given by

$$\overbrace{\inf_{K\in\mathcal{K}} \sup_{w\in\mathcal{H}_w} \frac{\|T_K(w)\|_z}{\|w\|_w}}^{\text{upper bound}} \geq \overbrace{\sup_{w\in\mathcal{H}_w} \inf_{K\in\mathcal{K}} \frac{\|T_K(w)\|_z}{\|w\|_w}}^{\text{lower bound}}.$$

The problem admits a so-called *soft-contrained* representation which consists in the following. Let γ^* be the attenuation level, which satisfies

$$\inf_{K\in\mathcal{K}} \sup_{w\in\mathcal{H}_w} \|T_K(w)\|_z^2 - {\gamma^*}^2\|w\|_w^2 \leq 0.$$

Define the parametrized cost (in $\gamma \geq 0$)

$$J_\gamma(K,w) := \|T_K(w)\|_z^2 - \gamma^2\|w\|_w^2.$$

Then, we wish to find the smallest value of $\gamma \geq 0$ under which the upper value is bounded (by zero) and this is called the soft-contrained game.

2.4 Examples

We next provide additional examples where we compute the loss ceiling $\overline{J}(A)$, the gain floor $\underline{J}(A)$, and the saddle-points (i^*, j^*) in pure strategies.

Example 2.3. In this first example, there is no saddle point as the existence condition in Theorem 2.1 is not matched.

$$\begin{array}{cc} & P_2 \\ P_1 & \begin{array}{|c|c|} \hline 6 & 0 \\ \hline -3 & 3 \\ \hline \end{array} \end{array}$$

To see this, observe that if player 1 plays top, then player 2 plays left and the payoff is 6 whereas if player 1 plays bottom, player 2 plays right and the payoff is 3. The conservative choice for player 1 is then bottom, and the loss ceiling is $\overline{J}(A) = a_{22} = 3$. For player 2 we have that, if he plays left, player 1 plays bottom, and if he plays right player 1 plays top. The conservative strategy for player 2 is then right and the gain floor is $\underline{J}(A) = a_{12} = 0$. The loss ceiling and gain floor do not coincide and the game admits no saddle-point.

Example 2.4. In this second example, the existence condition in Theorem 2.1 is matched and $(bottom, right)$ is a saddle-point.

$$P_1 \quad \overset{P_2}{\begin{array}{|c|c|} \hline -3 & 8 \\ \hline 4 & 4 \\ \hline \end{array}}$$

To see this, observe that if player 1 plays top, then player 2 plays right and the payoff is 8 whereas if player 1 plays bottom, player 2 plays equivalently left or right and the payoff is 4. The conservative choice for player 1 is then bottom, and the loss ceiling is $\overline{J}(A) = a_{21} = a_{22} = 4$. For player 2 we have that, if he plays left, player 1 plays top, and if he plays right player 1 plays bottom. The conservative strategy for player 2 is then right and the gain floor is $\underline{J}(A) = a_{22} = 4$ (except for indifferent unilateral deviation of players 2 from right to left). The loss ceiling and the gain floor coincide and the game admits a saddle-point, which is $(bottom, right)$.

Example 2.5. In this third example, there is no saddle-point as the existence condition in Theorem 2.1 is not matched.

$$P_1 \quad \overset{P_2}{\begin{array}{|c|c|} \hline -6 & 7 \\ \hline 2 & 1 \\ \hline \end{array}}$$

To see this, observe that if player 1 plays top, then player 2 plays right and the payoff is 7 whereas if player 1 plays bottom, player 2 plays left and the payoff is 2. The conservative choice for player 1 is then bottom, and the loss ceiling is $\overline{J}(A) = a_{21} = 2$. For player 2 we have that, if he plays left, player 1 plays top, and if he plays right player 1 plays bottom. The conservative strategy for player 2 is then right and the gain floor is $\underline{J}(A) = a_{22} = 1$. The loss ceiling and the gain floor do not coincide and the game admits no saddle-point.

Example 2.6. In this fourth and last example, the existence condition in Theorem 2.1 is matched and $(bottom, right)$ is a saddle-point.

$$P_1 \quad \overset{P_2}{\begin{array}{|c|c|} \hline -3 & 8 \\ \hline 2 & 4 \\ \hline \end{array}}$$

To see this, observe that if player 1 plays top, then player 2 plays right and the payoff is 8 and if player 1 plays bottom, player 2 plays still right and the payoff is 4 (right is a dominant strategy for player 2). The conservative choice for player 1 is then bottom, and the loss ceiling is $\overline{J}(A) = a_{22} = 4$. For player 2 we have that, if he plays left, player 1 plays top, and if he plays right player 1 plays bottom. The conservative strategy for player 2 is then right and the gain floor is $\underline{J}(A) = a_{22} = 4$. The loss ceiling and the gain floor coincide and the game admits a saddle-point which is $(bottom, right)$.

3

Computation of equilibria

This chapter is devoted to the computation of saddle-points and NE solutions. While the former can be obtained via linear programming, the latter need linear complementarity programming [Noam et al., 2007, Tijs, 2003]. Before introducing the general formulation of these mathematical programming problems, we provide graphical examples.

3.1 Example of graphical resolution

Let us compute the saddle-point in mixed strategies for the game

$$\begin{array}{cc} & P_2 \\ P_1 & \begin{array}{|c|c|} \hline 6 & 0 \\ \hline -3 & 3 \\ \hline \end{array} \end{array}$$

The mixed strategy for P_1 is $y^T = [y_1 \quad y_2] \in Y$ where

$$Y = \{y \in \mathbb{R}^2 : \sum_{i=1}^{2} y_i = 1,\, y_i \geq 0, \forall i = 1, 2\}.$$

Likewise, the mixed strategy for P_2 is $z^T = [z_1 \quad z_2] \in Z$ where

$$Z = \{z \in \mathbb{R}^2 : \sum_{j=1}^{2} z_j = 1,\, z_j \geq 0, \forall j = 1, 2\}.$$

The corresponding mean payoff is

$$J_m(A) = \sum_i \sum_j a_{ij} y_i z_j = y^T A z.$$

3.1.1 Conservative strategy for P_1

To compute the conservative strategy for P_1 let us consider the two pure actions of the opponent separately, then

$$\begin{cases} J_m(A) = y^T A z = 6y_1 - 3y_2, & z_1 = 1, \\ J_m(A) = y^T A z = 0y_1 + 3y_2, & z_2 = 1. \end{cases} \tag{3.1}$$

Figure 3.1 plots the mean payoff as a function of the mixed strategy of P_1 for the two cases of P_2 playing the left column ($z_1 = 1$) or right column ($z_2 = 1$).

Let us consider the point-wise maximum (worst payoff), which is given by

$$\max_z y^T A z = \max_{z_1, z_2} z_1(6y_1 - 3y_2) + z_2(0y_1 + 3y_2).$$

This is the boldface line in the figure. Let us compare the worst payoffs and look for the best worst-case, which yields

$$y^* = \arg\min_y \max_z y^T A z, \quad {y^*}^T = [0.49 \quad 0.51].$$

See the minimum of the boldface line. We reiterate the same calculation for P_2.

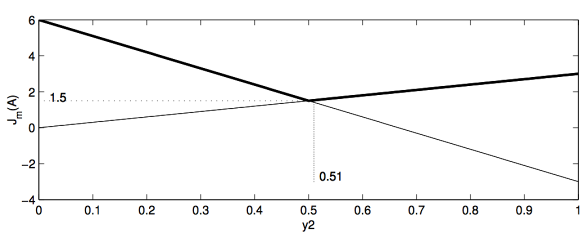

Figure 3.1: Graphical resolution for P_1.

3.1.2 Conservative strategy for P_2

The conservative strategy for P_2 is obtained by fixing the strategy for P_1 to one of the two alternative pure actions, so we have for the mean payoff

$$\begin{cases} J_m(A) = y^T Az = 6z_1 + 0z_2, & y_1 = 1, \\ J_m(A) = y^T Az = -3z_1 + 3z_2, & y_2 = 1. \end{cases} \tag{3.2}$$

Figure 3.2 plots the mean payoff as a function of the mixed strategy of P_2 for the two cases of P_1 playing the top row ($y_1 = 1$) or the down row ($y_2 = 1$).

The point-wise minimum (worst payoff), which is in boldface, is then obtained as

$$\min_y y^T Az = \min_{y_1,y_2} y_1(6z_1 + 0z_2) + y_2(-3z_1 + 3z_2).$$

The best worst-case is then given by (see the maximum of the boldface line)

$$z^* = \arg\max_z \min_y y^T Az, \quad {z^*}^T = [0.24 \quad 0.76].$$

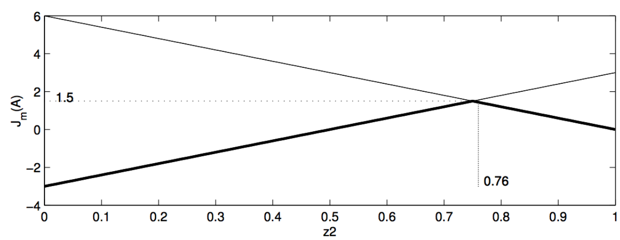

Figure 3.2: Graphical resolution for P_2.

Example 3.1. Let us compute the saddle point strategies for the following game

$$\begin{array}{cc} & P_2 \\ P_1 & \begin{array}{|c|c|} \hline -3 & 8 \\ \hline 4 & 4 \\ \hline \end{array} \end{array}$$

By repeating the same procedure as in the previous example we obtain the conservative strategies for player 1 and player 2 as depicted in

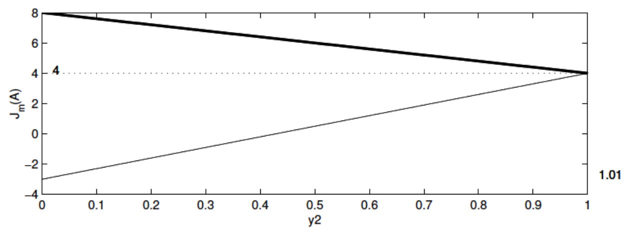

Figure 3.3: Graphical resolution for P_1: average payoff $J_m(A)$ (vertical axis) as a function of y_2 (horizontal axis).

Fig. 3.3 and 3.4 respectively. Note that the second column is weak dominant for P_2 from which we have that the point-wise maximum corresponds to the case $z_2 = 1$ (see the boldface line in Fig. 3.3) and is monotonic decreasing. Also note that the intersection between the two lines capturing the plot of the average payoff for $z_1 = 1$ and $z_2 = 1$ is at the extreme point $y_2 = 1$ and this is also the minimum of the point-wise maximum plot. The conservative strategy for P_2 leads to the two

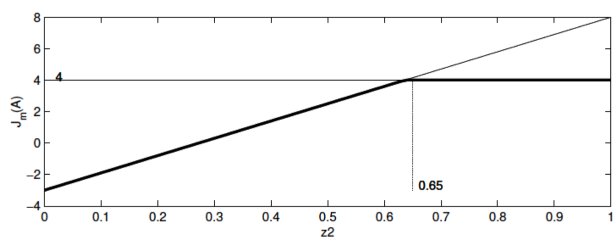

Figure 3.4: Graphical resolution for P_2: average payoff $J_m(A)$ (vertical axis) as a function of z_2 (horizontal axis).

lines in Fig. 3.4 illustrating the evolution of the average payoff in the two cases $y_1 = 1$ and $y_2 = 1$. The point-wise minimum is in boldface and returns infinite local maxima, i.e., all the points in the segment from $z_2 = 0.65$ to $z_2 = 1$. Therefore we have an infinite number of saddle-points.

Example 3.2. We repeat the same calculation for the following example

$$
\begin{array}{cc}
 & P_2 \\
P_1 & \begin{array}{|c|c|}\hline -3 & 8 \\ \hline 2 & 4 \\ \hline \end{array}
\end{array}
$$

We obtain the conservative strategies for player 1 and player 2 as depicted in Figs. 3.5 and 3.6 respectively. As in the previous example the second column is dominant (or more precisely it is strong dominant) for P_2 from which the point-wise maximum corresponds to the case $z_2 = 1$ (see the bold line in Fig. 3.5) and is monotonic decreasing. Then, we do not have any intersection between the two lines capturing the plot of the average payoff for $z_1 = 1$ and $z_2 = 1$ and the minimum of the point-wise maximum plot is again obtained for $y_2 = 1$.

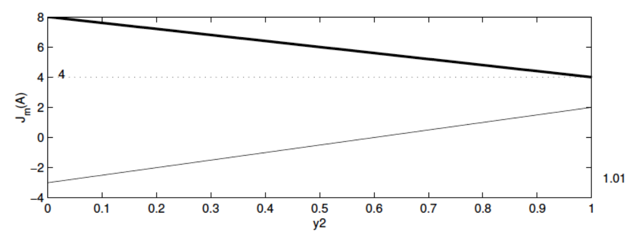

Figure 3.5: Graphical resolution for P_1: average payoff $J_m(A)$ (vertical axis) as a function of y_2 (horizontal axis).

The conservative strategy for P_2 leads to the two lines in Fig. 3.6 illustrating the evolution of the average payoff in the two cases $y_1 = 1$ and $y_2 = 1$. The point-wise minimum is in boldface and returns a unique maximum for $z_2 = 1$.

In the next section we generalize the treatment of saddle-points computation by introducing first a 3×4 matrix game and then providing the linear programming formulation in general terms.

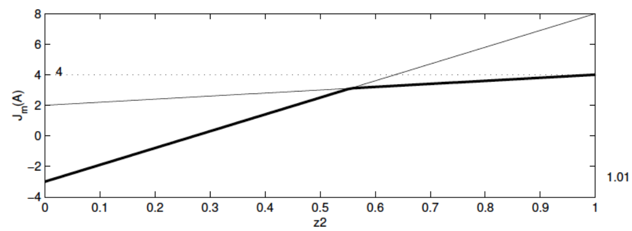

Figure 3.6: Graphical resolution for P_2: average payoff $J_m(A)$ (vertical axis) as a function of z_2 (horizontal axis).

3.2 Saddle-points via linear programming

Consider the game

$$
\begin{array}{cc}
 & P_2 \\
P_1 & \begin{array}{|c|c|c|c|} \hline 6 & 0 & 5 & 6 \\ \hline -3 & 3 & -4 & 3 \\ \hline 8 & 1 & 2 & 2 \\ \hline \end{array}
\end{array}
$$

The mixed strategy for P_1 is $y^T = [y_1 \quad y_2 \quad y_3] \in Y$ where

$$Y = \{y \in \mathbb{R}^3 : \sum_{i=1}^{3} y_i = 1,\ y_i \geq 0, \forall i = 1, \ldots, 3\}.$$

Analogously, the mixed strategy for P_2 is $z^T = [z_1 \quad z_2 \quad z_3 \quad z_4] \in Z$, where

$$Z = \{z \in \mathbb{R}^4 : \sum_{j=1}^{4} z_j = 1,\ z_j \geq 0, \forall j = 1, \ldots, 4\}.$$

In a saddle-point both players play conservative strategies, so we have

$$J_m(A) = \min_{y \in Y} \max_{z \in Z} y^T Az = \max_{z \in Z} \min_{y \in Y} y^T Az.$$

3.2.1 Linear program

The objective function for P_1 is $v_1(y) = \max_{z \in Z} y^T Az$, then P_1 solves

$$\min_{y \in Y} v_1(y).$$

It must hold

$$v_1(y) = \max_{z \in Z} y^T A z \geq y^T A z \quad \forall z \in Z,$$

which will constitute the constraints of the problem. After transposing we have

$$z^T A^T y \leq v_1(y) \Longrightarrow [z_1 \quad z_2 \quad z_3 \quad z_4] \begin{bmatrix} 6y_1 - 3y_2 + 8y_3 \\ 0y_1 + 3y_2 + 1y_3 \\ 5y_1 - 4y_2 + 2y_3 \\ 6y_1 + 3y_2 + 2y_3 \end{bmatrix} \leq v_1(y),$$

which must hold for every $z \in Z$. For the latter to be true it suffices that the above holds in each vertex of Z. Thus, by setting one component of z equal to one and the rest equal to zero, we obtain the following set of inequalities:

$$\begin{bmatrix} 6y_1 - 3y_2 + 8y_3 \\ 0y_1 + 3y_2 + 1y_3 \\ 5y_1 - 4y_2 + 2y_3 \\ 6y_1 + 3y_2 + 2y_3 \end{bmatrix} \leq \begin{bmatrix} v_1(y) \\ v_1(y) \\ v_1(y) \\ v_1(y) \end{bmatrix}.$$

Now, take $\tilde{y} = \frac{1}{v_1(y)} y$ (suppose $v_1(y) > 0$ without loss of generality) and rewrite the above set of inequalities as

$$\begin{bmatrix} 6\tilde{y}_1 - 3\tilde{y}_2 + 8\tilde{y}_3 \\ 0\tilde{y}_1 + 3\tilde{y}_2 + 1\tilde{y}_3 \\ 5\tilde{y}_1 - 4\tilde{y}_2 + 2\tilde{y}_3 \\ 6\tilde{y}_1 + 3\tilde{y}_2 + 2\tilde{y}_3 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

As $\tilde{y}_1 + \tilde{y}_2 + \tilde{y}_3 = \frac{1}{v_1(y)}$, minimizing $v_1(y)$ is equivalent to maximizing $\tilde{y}_1 + \tilde{y}_2 + \tilde{y}_3$. Then, we turn the latter problem into the following linear programming problem

$$\begin{aligned} &\max \tilde{y}_1 + \tilde{y}_2 + \tilde{y}_3 \\ &\begin{bmatrix} 6\tilde{y}_1 - 3\tilde{y}_2 + 8\tilde{y}_3 \\ 0\tilde{y}_1 + 3\tilde{y}_2 + 1\tilde{y}_3 \\ 5\tilde{y}_1 - 4\tilde{y}_2 + 2\tilde{y}_3 \\ 6\tilde{y}_1 + 3\tilde{y}_2 + 2\tilde{y}_3 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \\ &\tilde{y}_1 \geq 0, \tilde{y}_2 \geq 0, \tilde{y}_3 \geq 0. \end{aligned}$$

Analogously, by repeating the reasoning for P_2, we have the dual problem

$$\begin{array}{c} \min \tilde{z}_1 + \tilde{z}_2 + \tilde{z}_3 + \tilde{z}_4 \\ \begin{bmatrix} 6\tilde{z}_1 + 0\tilde{z}_2 + 5\tilde{z}_3 + 6\tilde{z}_4 \\ -3\tilde{z}_1 + 3\tilde{z}_2 - 4\tilde{z}_3 + 3\tilde{z}_4 \\ 8\tilde{z}_1 + 1\tilde{z}_2 + 2\tilde{z}_3 + 2\tilde{z}_4 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \\ \tilde{z}_1 \geq 0, \tilde{z}_2 \geq 0, \tilde{z}_3 \geq 0, \tilde{z}_4 \geq 0. \end{array}$$

In a nutshell, for the saddle point computation we need to solve the two linear programming problems

$$(P_1) \quad \begin{array}{rl} \max \mathbf{1}^T \tilde{y} & \\ A^T \tilde{y} & \leq \mathbf{1}, \\ \tilde{y} \geq 0. & \end{array} \qquad (P_2) \quad \begin{array}{rl} \min \mathbf{1}^T \tilde{z} & \\ A\tilde{z} & \geq \mathbf{1}, \\ \tilde{z} \geq 0. & \end{array} \tag{3.3}$$

The next section turns the attention to the computation of NE solutions via linear complementarity programming.

3.3 Nash equilibria via linear complementarity program

Suppose we wish to compute the NE solutions in mixed strategy for the "battle of the sexes"

	S	C
S	(2,1)	(0,0)
C	(0,0)	(1,2)

We know that the mixed NE is at the intersection of the best responses and as such is the action profile $y^* = [2/3, 1/3]$ and $z^* = [1/3, 2/3]$. Figure 3.7 depicts the best responses for the game at hand for player 1 (dashed) and player 2 (solid). In the horizontal axis we plot the action set for player 1 and in the vertical axis the one of player 2.

From the *Indifference Principle,* (introduced below Theorem 1.3 in Chap. 1) it holds that at an NE, playing mixed or changing to any strategy in the support is equivalent, namely,

$$u_1(S, z^*) = u_1(C, z^*) = u_1(y^*, z^*),$$

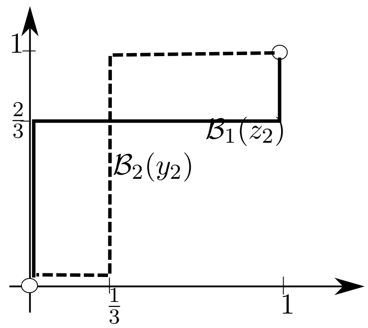

Figure 3.7: Best responses for the“Battle of the sexes”.

where S and C are the two strategies in the support of y^*. We revisit now two examples introduced in Chap. 1 before generalizing the treatment.

Example 3.3. Consider the Hawk and Dove or Chicken game, where $C = 6 > V = 4 > 0$:

	Hawk	Dove
Hawk	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	(V,0)
Dove	(0,V)	$\left(\frac{V}{2}, \frac{V}{2}\right)$

The best responses $\mathcal{B}_1(z_2)$ (solid) for P_1 and $\mathcal{B}_2(y_2)$ (dashed) for P_2 are displayed in Fig. 3.8. The NE is at the intersection between $\mathcal{B}_1(z_2)$ and $\mathcal{B}_2(y_2)$, namely, the action profile $y_2 = \frac{1}{3}$ and $z_2 = \frac{1}{3}$ for P_1 and P_2 respectively.

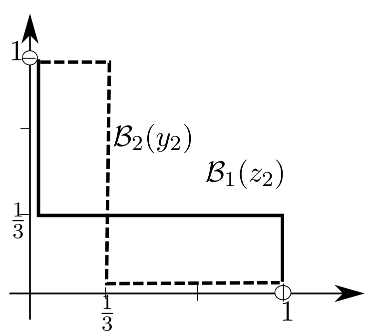

Figure 3.8: Best responses for the“Hawk and Dove” game.

Example 3.4. Stag-Hunt game:

	Stag	Hare
Stag	$\left(\frac{3}{2}, \frac{3}{2}\right)$	(0,1)
Hare	(1,0)	$\left(1, 1\right)$

The best responses $\mathcal{B}_1(z_2)$ (solid) for P_1 and $\mathcal{B}_2(y_2)$ (dashed) for P_2 are displayed in Fig. 3.9. The NE is at the intersection between $\mathcal{B}_1(z_2)$ and $\mathcal{B}_2(y_2)$, namely, the action profile $y_2 = \frac{1}{3}$ and $z_2 = \frac{1}{3}$) for P_1 and P_2 respectively.

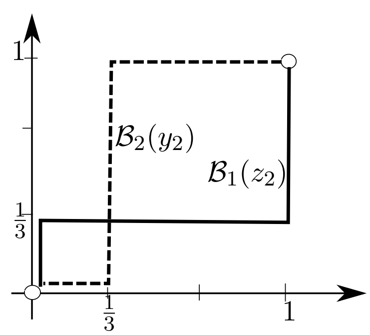

Figure 3.9: Best responses for the“Stag-Hunt” game.

Next we extend our discussion to any bimatrix game and derive the linear complementarity program in its abstract form.

3.3.1 Linear complementarity program

Generalizing, consider a bimatrix game (A, B), then for any NE (y^*, z^*) we have

$$\begin{array}{lll} {y^*}^T A z^* \geq y^T A z^*, \forall y & \Rightarrow & {y^*}^T A z^* \geq \sum_{j=1}^n a_{ij} z_j^*, \; \forall i = 1, \ldots, m, \\ {y^*}^T B z^* \geq {y^*}^T B z, \forall z & \Rightarrow & {y^*}^T B z^* \geq \sum_{i=1}^m b_{ij} y_i^*, \; \forall j = 1, \ldots, n. \end{array}$$

From the indifference principle we then obtain

$$y_i^* > 0 \Rightarrow \sum_{j=1}^n a_{ij} z_j^* = {y^*}^T A z^*; \quad z_j^* > 0 \Rightarrow \sum_{i=1}^m b_{ij} y_i^* = {y^*}^T B z^*.$$

Now, let us take $u_j = \frac{z_j^*}{{y^*}^T A z^*}$ and $v_i = \frac{y_i^*}{{y^*}^T B z^*}$, then we have

$$\begin{cases} \sum_{j=1}^n a_{ij} u_j \le 1, \forall i = 1, \dots, m, \\ \sum_{i=1}^m b_{ij} v_i \le 1, \; \forall j = 1, \dots, n, \\ \forall i \, v_i > 0 \quad \Rightarrow \quad \sum_{j=1}^n a_{ij} u_j = 1, \\ \forall j \, u_j > 0 \quad \Rightarrow \quad \sum_{i=1}^m b_{ij} v_i = 1. \end{cases} \tag{3.4}$$

Let us introduce the slack variables $r \in \mathbb{R}^m$ and $t \in \mathbb{R}^n$ and rewrite the problem above as a *linear complementarity program*:

$$\begin{cases} Au + r = 1, \\ B^T v + t = 1, \\ v^T r = 0, \\ u^T t = 0, \\ r \ge 0, \, t \ge 0. \end{cases} \quad \text{or} \quad \begin{cases} \overbrace{\begin{bmatrix} 0 & A \\ B^T & 0 \end{bmatrix}}^{H} \overbrace{\begin{bmatrix} v \\ u \end{bmatrix}}^{x} + \overbrace{\begin{bmatrix} r \\ t \end{bmatrix}}^{s} = 1 \\ \begin{bmatrix} v^T & u^T \end{bmatrix} \begin{bmatrix} r \\ t \end{bmatrix} = 0, \; \begin{bmatrix} r \\ t \end{bmatrix} \ge 0. \end{cases} \tag{3.5}$$

From $s := [r \quad t]^T \ge 0$, the above problem is equivalent to

$$\begin{cases} 1 - Hx \ge 0, \\ x^T(1 - Hx) = 0, \\ x \ge 0, \end{cases} \text{ solved as } \begin{cases} \min_x x^T(1 - Hx) \\ 1 - Hx \ge 0, \\ x \ge 0, \end{cases} \qquad \text{(bilinear prog.)}$$

which is in the form of a quadratic bilinear program and this concludes our derivation.

4

Refinement on equilibria, Stackelberg equilibrium and Pareto optimality

In this chapter, first we elaborate on properties of NE solutions such as *payoff dominance*, *risk dominance*, or *subgame perfectness*. In the second part, we introduce the Stackelberg equilibrium and discuss Pareto optimality.

4.1 Refinement on Nash Equilibrium solutions

NE solutions may enjoy several different properties, which can make some equilibria more interesting than others in terms of their impact on the global welfare of the system. This is particularly important in the design of incentives that may enforce the agents or players to converge to certain equilibria rather than others. Let us look at two main characteristics of NE solutions, payoff dominance and risk dominance.

4.1.1 Admissible (Payoff dominant) NE

Admissible NE solutions, also called payoff dominant equilibria, are associated to a higher payoff for all the players.

Definition 4.1. An NE strategy pair is *admissible* or *payoff dominant* if there exists no better (for all players) NE strategy pair.

The concept is illustrated in the following example.

Example 4.1. Let the two-player game in Fig. 4.1 be given. We observe that the action profile (R, R) is payoff dominant as both players get more than what they would get if they played one of the other two NE solutions, namely, (L, L) or (L, M). Actually in (R, R) both players get 1 which exceeds what both players get in (L, L), that is 0 and also what both players get in (L, M), namely -1 for player 1 and 0 for player 2.

	L	**M**	**R**
L	(0,0)*	(-1,0)*	(-3,-1)
R	(-2,1)	(-2,0)	(1,1)*

Admissible NE

Figure 4.1: Payoff dominance.

4.1.2 Risk dominant NE

The second property we wish to discuss here is risk dominance. Consider the parametrized coordination game, where $A > B$, $D > C$ for player 1, and $a > b$, $d > c$ for player 2:

	H	G
H	(A, a)	(C, b)
G	(B, c)	(D, d)

The formal definition is as follows.

Definition 4.2. (Harsanyi and Selten, 1988, Lemma 5.4.4, [Harsanyi and Selten, 1988]) An NE strategy pair (G, G) risk dominates the other NE strategy pair (H, H) if the product of the deviation losses is highest for (G, G), namely if the following inequality holds:

$$(C - D)(c - d) \geq (B - A)(b - a).$$

A largely accepted belief, supported by evolutionary biologists, is that while payoff dominance is in principle more convenient to all the players and as such should arise naturally, in most experimental situation in nature players (animals) usually converge to risk dominance equilibria [Kandori et al., 1993, Young, 1993]. The classical Stag-Hunt game explains further the difference between risk and payoff dominance.

Example 4.2. In the Stag-Hunt game $(Hare, Hare)$ is risk dominant, while $(Stag, Stag)$ is payoff dominant

	Stag	Hare
Stag	$\left(\frac{3}{2}, \frac{3}{2}\right)$	(0,1)
Hare	(1,0)	$(1, 1)$

Note that if the opponent plays with probabilities $(1/2, 1/2)$, then the action $Hare$ has a higher expected payoff.

4.1.3 Subgame perfect NE

For extensive games, where there is an explicit order of events, one can further distinguish so-called subgame perfect NE solutions. This concept builds upon the notion of subgame as provided below.

Definition 4.3. A subgame is a subset of nodes that still forms a game.

Given a subgame, a subgame perfect NE must be still an NE for the residual game represented by the subgame.

Definition 4.4. An equilibrium is subgame perfect if when played from any point in the game it is a Nash equilibrium.

The computation of subgame perfect NE is based on dynamic programming as shown in the examples below.

Example 4.3. Figure 4.2 depicts a two-player extensive game (left) and its normal form representation (right). At stage 2 P_2's rational choice is (D) (red line). By backwards induction at stage 1 P_1's best response is B (thus he gets 5 rather than 4 if he played A). From the normal form representation (A, C) is also an NE as deviating to B is not beneficial to P_1 neither is deviating to D to P_2. However, such an NE is not subgame perfect.

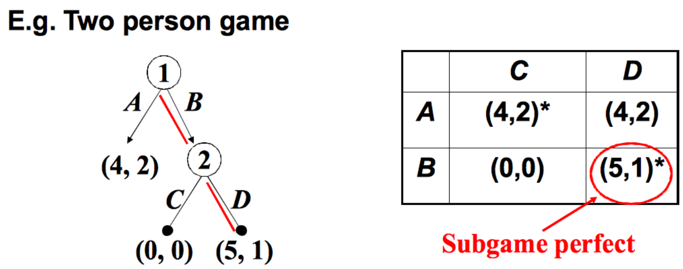

Figure 4.2: Subgame perfect NE based on dynamic programming.

Example 4.4. Figure 4.3 depicts a three stage two-player extensive game (left) and its normal form representation (right). The game has four different states. At stage 3, state 4, P_1 selects R_4 (red line). By backwards induction, at stage 2 P_2's rational choice is L_2 in state 2 and is equivalently L_3 (red line) or R_3 in state 3 (he would get 5 in both cases). At stage 1, P_1's best response to L_3 is L_1 (red), while the best response to R_3 is R_1 (blue line). Both (L_1R_4, L_2L_3) and (R_1R_4, L_2R_3) are subgame perfect NE solutions. From the normal form representation we have several other NE solutions which are not subgame perfect.

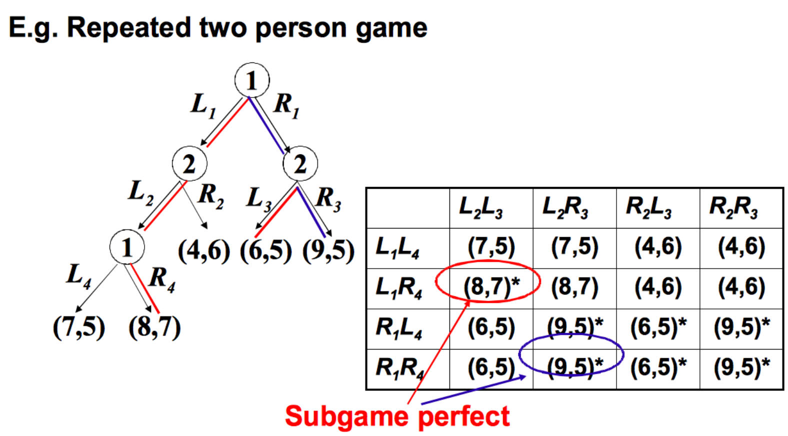

Figure 4.3: Subgame perfect NE based on dynamic programming.

4.2 Stackelberg equilibrium

The Stackelberg equilibrium [von Stackelberg, 1934], differently from the Nash equilibrium, assumes that the game admits a hierarchical structure. In particular, suppose that a leader announces and enforces his best strategy by taking into account the rational reaction of the followers. To introduce a formal definition of Stackelberg equilibrium it is useful to recall the definition of best response set, which is reiterated below

$$\mathcal{B}_i(a_{-i}) := \{a_i^* \in \mathcal{A}_i |\, u_i(a_i^*, a_{-i}) = \max_{a_i \in \mathcal{A}_i} u_i(a_i, a_{-i})\}.$$

We are then in the position to define a Stackelberg equilibrium as follows.

Definition 4.5. An action profile (a_1^S, a_2^S) is a Stackelberg equilibrium (SE) for player 1 if $a_2^S \in \mathcal{B}_2(a_1^S)$ and

$$u_1(a_1^S, a_2^S) \geq u_1(a_1, a_2), \quad \forall a_1 \in \mathcal{A}_1,\, a_2 \in \mathcal{B}_2(a_1).$$

We provide next multiple examples expanding on the relation between NE solutions and Stackelberg equilibria. We will see that in certain cases the Stackelberg equilibrium can also be an NE. However, in general, a Stackelberg equilibrium differs from an NE, and may yield better payoffs for all the players or for only a subset of players.

Example 4.5. In the Prisoner's dilemma, if the leader is P_1, then he knows that if he picks C, P_2 (follower) would respond with D (red circle) and he would get 0, whereas by picking D as P_2 would still respond with D (red circle) he would get 1, then his rational choice is D. The Stackelberg equilibrium SE_1 when P_1 is leader is then (D, D). Analogously, if the leader is P_2, then he knows that if he picks C, P_1 (follower) would respond with D (blue circle) and he would get 0, whereas by picking D as P_1 would still respond with D (blue circle) he would get 1, then his rational choice is again D. The Stackelberg equilibrium SE_2 when P_2 is leader is again (D, D). In this example both Stackelberg equilibria are the same and also coincide with the unique pure NE of the game.

	C	D
C	(3,3)	(0,4)
D	(4,0)	(1,1)*

SE_1

SE_2

Figure 4.4: Stackelberg equilibrium for the Prisoners' dilemma.

Example 4.6. In the example of Fig. 4.5, if the leader is P_1, then he knows that i) if he picks L, P_2 (follower) would respond with L (red circle) and he would get 0, ii) if he picks M, P_2 (follower) would respond with M (red circle) and he would get -1 whereas iii) by picking R as P_2 would respond with R (red circle) he would get -2, then his rational choice is L. The Stackelberg equilibrium SE_1 when P_1 is leader is then (L, L). Analogously, if the leader is P_2, then he knows that i) if he picks L, P_1 (follower) would respond with R (blue circle) and he would get 0, ii) if he picks L, P_1 (follower) would respond with M (blue circle) and he would get 0, whereas by picking R as P_1 would respond with L (blue circle) he would get $\frac{2}{3}$, then his rational choice is R. The Stackelberg equilibrium SE_2 when P_2 is leader is again (L, R). In this example, SE_1 is better than the NE for both players, and SE_2 is better than the NE only for P_2 (leader). The NE is not a Stackelberg equilibrium.

	L	M	R
L	(0,1)	(-2,-1)	(-3/2,2/3)
M	(-1,-2)	(-1,0)*	(-3,-1)
R	(1,0)	(-2,-1)	(-2,1/2)

SE_1

SE_2

Figure 4.5: Example of Stackelberg equilibrium.

Example 4.7. Consider the infinite game in Fig. 4.6. Player 1 plays the horizontal coordinate, while player 2 plays the vertical coordinate. The global maximum is P for player 1, and Q for player 2. The best response of P_2 to P_1 is the curve $\beta(a)$ (dash-dot). The Stackelberg equilibrium when P_1 is leader is (a_s, b_s) (this point is crossed by the tangent to P_1's level curves).

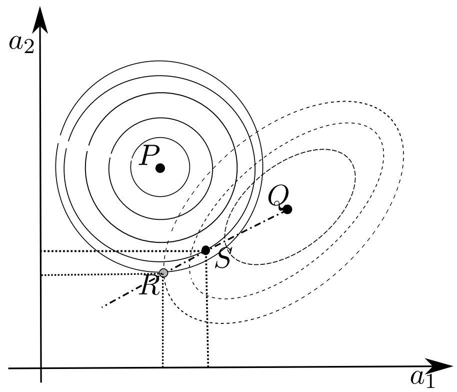

Figure 4.6: Two-player continuous infinite game. Level curves of player 1 (solid) and player 2 (dashed), action space of player 1 (horizontal axis), and of player 2 (vertical axis). Global maximum is P for player 1 and Q for player 2 while the NE is point R and the Stackelberg equilibrium is point S. Courtesy by Alberto Bressan, Noncooperative Differential Games. A Tutorial (2010) [Bressan, 2010].

4.2.1 Non-uniqueness

As for the NE, also the Stackelberg equilibrium may not be unique. In such a case, one can always introduce additional criteria to refine on Stackelberg equilibria. For instance, one could seek to minimize the risk, and play conservative. This is illustrated in detail in the following example.

Example 4.8. In the example of Fig. 4.7, when the leader is P_1, he knows that i) if he picks L, P_2 (follower) would respond indifferently with L or M (red circle) and he would get 0 or -1 respectively; ii) if he picks R, P_2 (follower) would respond indifferently with L or R (red circle) and he would get -2 or 1 respectively. Note that the expected payoff by playing L or R for P_1 is always $-\frac{1}{2}$. However, the worst payoff is better in the first case (L), then minimizing the risk would yield (L, M) as SE_1. Also note that SE_1 is worse than the NE for both players.

E.g. Bimatrix Game

SE

	L	M	R
L	(0,0)*	(-1,0)*	(-3,-1)
R	(-2,1)	(-2,0)	(1,1)*

Admissible NE

Figure 4.7: Nonunique Stackelberg equilibrium and risk-minimization.

4.3 Pareto optimality

We conclude this chapter introducing a property which may or may not be enjoyed by equilibria, that is, Pareto optimality. It must be said that, equilibria that are also Pareto optimal represent extremely stable solutions in that not only no player is better off by changing actions, but also no players can be better off by jointly deviating without causing a loss for at least one player.

Definition 4.6. A pair of strategies (a_1^{PO}, a_2^{PO}) is said to be Pareto optimal (PO) if there exists no other pair (a_1, a_2) such that for $i = 1, 2$

$$u_i(a_1, a_2) > u_i(a_1^{PO}, a_2^{PO}) \text{ and } u_{-i}(a_1, a_2) \geq u_{-i}(a_1^{PO}, a_2^{PO}).$$

In other words, given a PO solution, it is not possible to strictly increase the payoff of one player without strictly decreasing the payoff of the other.

Example 4.9. An interesting question is then: what solutions are PO in the Prisoner's Dilemma? Actually, we observe three different PO solutions, that is (C, D), (D, C) and (C, C). Indeed, for each of the these action profiles, any deviation causes a loss for at least one of the players.

	C	D
C	(3,3)	(0,4)
D	(4,0)	(1,1)

Figure 4.8: In the Prisoner's dilemma (C, D), (D, C) and (C, C) are all Pareto optimal solutions.

In the rest of this chapter we analyze Stackelberg equilibria and Pareto optimal solutions for classical games like the Battle of the sexes, the Coordination game, and the Hawk and Dove game.

Example 4.10. (Battle of sexes) When player 1 is the leader the Stackelberg equilibrium is (S, S) whereas when player 2 is the leader the Stackelberg equilibrium is (C, C). The two Stackelberg equilibria are also the only Pareto optimal solutions of the game.

	S	C
S	$(2,1)$	$(0,0)$
C	$(0,0)$	$(1,2)$

Example 4.11. (Coordination game) In this example the Stackelberg equilibrium is $(Mozart, Mozart)$ independently of the leader. The Stackelberg equilibrium is also the only Pareto optimal solution of the game.

	$Mozart$	$Mahler$
$Mozart$	(2,2)	(0,0)
$Mahler$	(0,0)	(1,1)

Example 4.12. (Hawk and dove game) Suppose that the cost of fight exceeds the prize of victory, that is $C > V > 0$ in the bimatrix game below.

	$Hawk$	$Dove$
$Hawk$	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
$Dove$	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

When player 1 is the leader the Stackelberg equilibrium is $(Hawk, Dove)$ whereas when player 2 is the leader the Stackelberg equilibrium is $(Dove, Hawk)$. All solutions are Pareto optimal except $(Hawk, Hawk)$.

5

Cooperative game theory

This chapter provides an introduction to coalitional games with transferable utilities (TU games). After presenting some examples borrowed from operational research such as the minimum spanning tree game, the permutation game, and the max-flow game, we review solution concepts like the imputation set. The last part of the chapter, following [Yeung and Petrosjan, 2006], reframes the same topics within a stochastic dynamic scenario thus covering what is known in the literature as *cooperative differential games*. For these games we introduce the notion of *dynamic stability or time consistency*.

5.1 Coalitional games with transferable utilities (TU games)

A TU game is a tuple $\langle N, v \rangle$, where $N = \{1, \dots, n\}$ is the set of players and $v : 2^N \to \mathbb{R}$ is the *characteristic function*. Thus, $v(S)$ is the value of coalition $S \subseteq 2^N$. Such a value is computed assuming that all the members not in the coalitions will play their joint worst-case actions. This turns the computation into a min-max optimization problem. Indeed, the value is obtained as the minimum over the set of joint actions of all the players belonging to the coalition, and as the maximum over the set of joint actions of all the players outside the coalition. In a more compact way we say that the value of a coalition is the total payoff that the members of the coalition can guarantee for themselves at least.

Example 5.1. A TU game reformulation of the Prisoner's dilemma is obtained by considering as value of the coalition the number of years of freedom that the members of the coalitions can guarantee for themselves against any possible play of the members outside the coalition. Thus every single player can guarantee 1 year of freedom in the worst-case, while both players cooperating can guarantee for themselves 6 years of freedom in total. This gives: $N = \{1, 2\}$,

$$v(\{1\}) = v(\{2\}) = 1, \quad v(\{1, 2\}) = 6.$$

	C	D
C	(3,3)	(0,4)
D	(4,0)	(1,1)

Figure 5.1: The Prisoners' dilemma as a TU game.

Example 5.2. For the two-player extensive game in Fig. 5.2, the TU reformulation is given by: $N = \{1, 2\}$,

$$v(\{1\}) = 4, \quad v(\{2\}) = 2, \quad v(\{1, 2\}) = 12.$$

Indeed $v(S)$ is how much each coalition can guarantee for itself at least. P_1 alone can play R and get 4 in the worst-case (when P_2 plays l_2). P_2 by playing alone can get 2 by choosing r_1 in state 1, and can get 8 by playing l_2 in state 2. The state depends on the P_1's action, so in the worst-case P_2 can get 2. Both players in coalition can get a maximal total payoff of 12 by playing the joint actions (R, l_2).

Example 5.3. For the three-player extensive game in Fig. 5.3, the TU reformulation is given by: $N = \{1, 2, 3\}$,

$$\begin{array}{rrr} v(\{1\}) = 10, & v(\{2\}) = 0, & v(\{3\}) = 0 \\ v(\{1, 2\}) = 14 & v(\{1, 3\}) = 11 & v(\{2, 3\}) = 0 \end{array}$$
$$v(\{1, 2, 3\}) = 16.$$

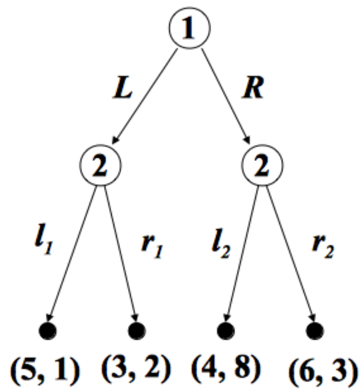

Figure 5.2: Two person extensive game as a TU game.

Indeed, P_1 alone can play M_1 and get 10 and the game terminates (in which case P_2 and P_3 actions do not play any role). For the same reason P_2 and P_3 by playing alone can guarantee for themselves no more than 0. Same result even if P_2 and P_3 play in coalitions. P_1 and P_2 in coalition can agree on P_1 playing R_1 which guarantee a total of $8+6=14$. P_1 and P_3 in coalition can agree on the joint action (R_1, R_3) which yields a total of $11+0=11$. All players together (grand coalition) can get a total of $6+8+2=16$ [Tijs, 2003].

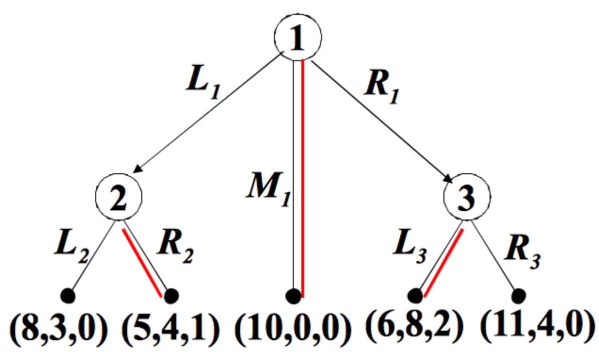

Figure 5.3: Three person extensive game as a TU game.

5.2 Operational research games

Operational research games [Tijs, 2003] are classical operational research problems revisited in the case where we have multiple interacting decision makers, which we now call players. When all the players agree in playing in coalition as if they were a unique entity then such games coincide with the classical optimization problems covered in operational research textbooks [Hillier and Lieberman, 2001].

5.2.1 Minimum spanning tree game

Communities 1, 2, and 3 want to be connected to a nearby power source. The possible transmission links and corresponding costs are as in Fig. 5.4. Every community can connect directly to the source paying 100, 90, and 80 respectively. Communities 1 and 2 can connect at the minimum total cost of $90 + 30 = 130$ through the tree $\{(source, 2), (2, 1)\}$. Similarly communities 1 and 3 can connect at the minimum total cost of $80 + 30 = 110$ through the tree $\{(source, 3), (3, 1)\}$. Communities 2 and 3 can connect at the minimum total cost of $80 + 30 = 110$ through the tree $\{(source, 3), (3, 2)\}$. All communities can connect at the minimum total cost of $80 + 30 + 30 = 140$ through the spanning tree $\{(source, 3), (3, 1), (3, 2)\}$. The cost game $\langle N, c\rangle$ is then given by: $N = \{1, 2, 3\}$,

$$\begin{gathered} c(\{1\}) = 100, \qquad c(\{2\}) = 90, \qquad c(\{3\}) = 80, \\ c(\{1,2\}) = 130, \quad c(\{1,3\}) = 110, \quad c(\{2,3\}) = 110, \\ c(\{1,2,3\}) = 140. \end{gathered}$$

The value of the coalition is the money saved by all the participants when they play together instead of alone. Then, we have

$$v(S) = \sum_{i \in S} c(\{i\}) - c(S).$$

The corresponding *cost saving game* $\langle N, v\rangle$: $N = \{1, 2, 3\}$, is then given by

$$v(\{1\}) = 0, \qquad v(\{2\}) = 0, \qquad v(\{3\}) = 0,$$

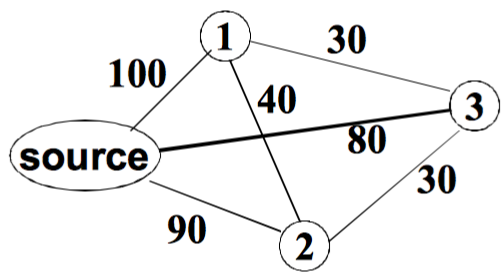

Figure 5.4: Minimum spanning tree problem as TU game.

$$v(\{1,2\}) = 60, \qquad v(\{1,3\}) = 70, \qquad v(\{2,3\}) = 60,$$
$$v(\{1,2,3\}) = 130.$$

It must be observed that in the case where all the players agree to play in the grand coalition the game takes on the form of a classical minimum spanning tree problem [Hillier and Lieberman, 2001].

5.2.2 Permutation game

There are n players $i = 1, 2, \ldots, n$; each one possesses a machine M_i and a job J_i to be processed. Any machine M_j can process any job J_i, but no machine is allowed to process more than one job. Coalition formation and side payments are allowed. If a person does not cooperate, his job has to be processed on his own machine. The cost of processing job J_i on machine M_j equals k_{ij}, where for each coalition $S \in 2^N \setminus \emptyset$,

$$c(S) = \min_{\sigma} \sum_{i \in S} k_{i\sigma(i)} \quad \sigma \text{ is any permutation of } 1, \ldots, n.$$

Consider for instance a three person permutation game whose cost matrix is

$$\begin{array}{cc} \text{Machine} & \\ \begin{bmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \\ 4 & 8 & 12 \end{bmatrix} & \text{Job} \end{array} \tag{5.1}$$

The corresponding costs and coalitions' values are summarized in the table below.

S	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$c(S)$	0	1	6	12	5	7	17	13
$v(S)$	0	0	0	0	2	6	1	6

Note that when all players form a grand coalition, the problem turns into a standard assignment problem [Hillier and Lieberman, 2001].

5.2.3 Max-flow game

Let us consider the flow from source to sink in the example of Fig. 5.5. Each arc has an owner and some capacity constraints. In particular, arc 1 has capacity 4 and the owner is l_1; arc 2 has capacity 5 and the owner is l_2; arc 3 has capacity 10 and the owner is l_3. The value of a coalition is the maximum flow through the network from source to sink, where one uses only arcs which are owned by the members of the coalition S. The TU game is then given by $N = \{1,2,3\}$,

$$\begin{aligned} v(\{1\}) &= 0, \quad & v(\{2\}) &= 0, \quad & v(\{3\}) &= 0, \\ v(\{1,2\}) &= 0, \quad & v(\{1,3\}) &= 4, \quad & v(\{2,3\}) &= 5, \\ & & v(\{1,2,3\}) &= 9. & & \end{aligned}$$

The game looks much alike a max-flow optimization problem [Hillier

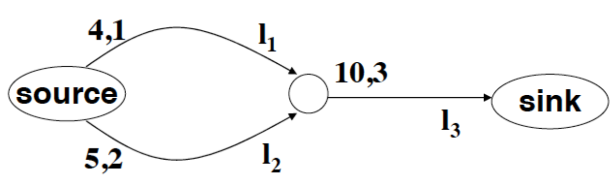

Figure 5.5: Max-flow problem as TU game: the labels $4,1$ and l_1 on one of the arcs mean that this is arc 1 with maximum capacity equal to 4 and whose owner is l_1.

and Lieberman, 2001] when all players form a grand coalition and act as if there were just one single decision maker. With in mind the formulation of TU games provided above, we now turn to look at a first solution concept, the imputation set.

5.3 Imputation set

Given a TU game a main question is how to divide the costs or rewards among the participants to the coalition. A partial answer to this question is given by the imputation set. The imputation set $I(v)$ is a convex polyhedral set that contains all allocations that are

- *efficient or Pareto optimal*, that is, all the components sum up to the value of the grand coalition, and
- *individual rational*, namely no individual benefits from splitting from the grand coalition and playing alone.

More formally, the imputation set is defined as

$$I(v) = \{x \in \mathbb{R}^n | \overbrace{\sum_{i \in N} x_i = v(N)}^{\text{Efficiency}}, \; \underbrace{x_i \geq v(\{i\}), \forall i \in N}_{\text{individual rationality}} \}.$$

The imputation set may in general be empty, which implies that, for every efficient allocation, there is always at least one player who is better off by quitting the grand coalition. A necessary and sufficient condition for the imputation set to be nonempty is that the sum of the values of the single players does not exceed the value of the grand coalition:

$$I(v) \neq \emptyset \quad \text{ iff } v(N) \geq \sum_{i \in N} v(\{i\}).$$

A main fact that can help in the computation of the imputation set is that the imputation set $I(v)$ is the convex hull of the points $f^1, f^2, \ldots, f^n$ where

$$f_k^i = \begin{cases} v(N) - \sum_{k \in N \setminus \{i\}} v(\{k\}) & k = i \\ v(\{k\}) & k \neq i. \end{cases} \tag{5.2}$$

Note that the generic vector f^i can be interpreted as a tentative allocation recommended by player i and obtained as follows. Given that f_k^i is the revenue that player i is willing to allocate to player k, then player i allocates to any other player the amount of his exact value, that is $f^i_k = v(\{k\})$, for $k \neq i$, and takes all the rest for himself, i.e., $f^i_i = v(N) - \sum_{k \in N \setminus \{i\}} v(\{k\})$. This computational concept is further illustrated in the next example.

Example 5.4. Let $\langle N, v \rangle$ be a three person game with

$$v(\{1\}) = v(\{3\}) = 0, \quad v(\{2\}) = 3, \quad v(\{1,2,3\}) = 5.$$

The imputation set $I(v)$ is the triangle with vertices $f^1 = (2,3,0)$, $f^2 = (0,5,0)$, and $f^3 = (0,3,2)$ as displayed in Fig. 5.6. Actually,

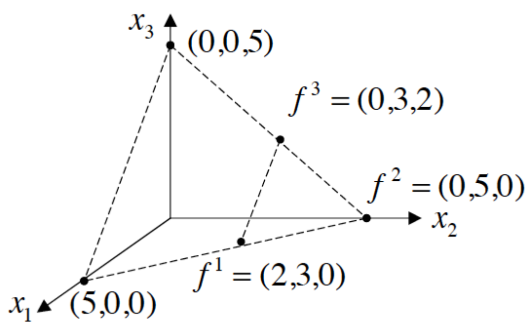

Figure 5.6: Example of imputation set.

player 1 says "I will give player 2 and 3 each one exactly their own value, which is 3 and 0 respectively, and will keep the rest for me, namely 2" from which $f^1 = (2,3,0)$. We can repeat the same reasoning for player 2 and 3 and obtain $f^2 = (0,5,0)$, and $f^3 = (0,3,2)$.

The following classification helps to immediately recognize the existence of stable allocation rules, namely allocation rules in the imputation set.

5.4 Properties

TU games divide into the following categories. Superadditive games are those which satisfy

$$v(S \cup T) > v(S) + v(T), \, \forall S, T \in 2^N : \, S \cap T = \emptyset.$$

Subadditive games are those characterized by

$$v(S \cup T) < v(S) + v(T),\ \forall S, T \in 2^N : S \cap T = \emptyset.$$

A third category is the one of additive games for which it holds

$$v(S \cup T) = v(S) + v(T),\ \forall S, T \in 2^N : S \cap T = \emptyset.$$

These are also called *inessential games* as how to divide earnings in an additive game is a trivial matter. Every player can get exactly his own value.

5.5 Cooperative differential games

We now turn to discuss the same topics illustrated above but in a dynamic and stochastic scenario [Yeung and Petrosjan, 2006]. To this purpose, consider the state dynamics below, where u_i is the control of player $i \in N$:

$$\dot{x}(s) = f[s, x(s), u_1(s), u_2(s), \ldots, u_n(s)], \quad x(t_0) = x_0.$$

We assume that the payoff of the generic player $i \in N$ is given by the following finite horizon integral

$$\int_{t_0}^{T} g^i[s, x(s), u_1(s), u_2(s), \ldots, u_n(s)]ds + q^i(x(T)).$$

In the cooperative game starting in state x_0 at time t_0, which we compactly denote by $\Gamma_v(x_0, T - t_0)$, players agree on joint controls $u_1^*(s), \ldots, u_n^*(s)$ that maximize the total payoff

$$v(N; x_0, T - t_0) = \sum_{i \in N} \int_{t_0}^{T} g^i[s, x(s), u_1^*(s), \ldots, u_n^*(s)]ds + q^i(x(T)),$$

and on a mechanism to distribute the total payoff among players. Then, the corresponding game $\Gamma_v(x_0, T - t_0)$ is called a cooperative differential game in characteristic function form.

Denote by $\xi_i(x_0, T - t_0)$ the share of the players $i \in N$ from the total payoff $v(N; x_0, T - t_0)$.

Then the value of a single player $i \in N$, $v(\{i\}; x_0, T - t_0)$, is how much he can guarantee for himself when all others play against, which is formally given by

$$\max_{u_i} \min_{u_j, j \neq i} \int_{t_0}^{T} g^i[s, x(s), u_1(s), u_2(s), \ldots, u_n(s)]ds + q^i(x(T)).$$

Adapting the notion of imputation set to the current scenario we arrive at the following definition.

Definition 5.1. A vector of shares

$$\xi(x_0, T - t_0) = [\xi_1(x_0, T - t_0) \ldots \xi_n(x_0, T - t_0)]$$

is called *imputation* if

$$\begin{array}{lll} (i) & \xi_i(x_0, T - t_0) \geq v(\{i\}; x_0, T - t_0), \quad \forall\, i \in N & \text{(rational)} \\ (ii) & \sum_{j \in N} \xi_j(x_0, T - t_0) = v(N; x_0, T - t_0) & \text{(efficient).} \end{array} \tag{5.3}$$

Now, consider the family of games along the optimal trajectory $\{x^*(s)\}_{s=t_0}^{T}$:

$$\{\Gamma_v(x^*(t), T - t), \quad t_0 \leq t \leq T\}.$$

Let us split the total share ξ_i into current ω_i and future share η_i:

$$\begin{aligned} &\eta_i[\xi_i(x_0, T - t_0); x^*(t), T - t] \\ &= \xi_i(x_0, T - t_0) - \omega_i[\xi_i(x_0, T - t_0); x^*(\cdot), t - t_0]. \end{aligned} \tag{5.4}$$

Definition 5.2. An imputation

$$\xi(x_0, T - t_0) = [\xi_1(x_0, T - t_0) \ldots \xi_n(x_0, T - t_0)]$$

is said *dynamically stable or time consistent* if

$$\begin{array}{ll} (i) & \eta_i[\xi_i(x_0, T - t_0); x^*(t), T - t] \geq v(\{i\}; x^*(t), T - t),\ \forall\, i \in N, \\ (ii) & \sum_{j \in N} \eta_j[\xi_j(x_0, T - t_0); x^*(t), T - t] = v(N; x_0, T - t). \end{array}$$

Essentially we are defining a stream of infinite TU games, one for each time t in the interval $[t_0, T)$, and for each of such games we look at allocations in the corresponding imputation set. We refer the reader to [Engwerda, 2005] and [Reddy and Engwerda, 2013] for a characterization of the full Pareto curve. Alternative formulations of robust dynamic cooperative games are also provided in [Bauso and Timmer, 2009] and [Bauso and Timmer, 2012]. Connections between consensus problems and distributed allocation in dynamics cooperative games are explored in [Bauso and Nedić, 2013]. Social optimal equilibria in multi-inventory applications are explored in [Bauso et al., 2008, 2009]. Let us proceed with a look at other seminal solution concepts such as core, Shapley value, and nucleolus.

6

Core, Shapley value, nucleolus

Given a TU game as introduced in the previous chapter, we now elaborate further on solution concepts yielding stable allocation rules. In particular we discuss the core, the Shapley value, and the nucleolus. In the last part of this chapter we also discuss computational aspects of the nucleolus.

6.1 Core

A solution concept which is largely well-known among economists is the one of *Core* of a TU game. In addition to the imputation set, the core also provides allocations that are stable with respect to any sub-coalitions. The core strengthens the conditions valid for the imputation set in that not only do players not benefit from quitting the grand coalition and playing alone, but also players do not benefit from creating any sub-coalition. Thus the core is still a polyhedral set which is included in the imputation set.

Definition 6.1. The core of a game $\langle N, v\rangle$ is the set of allocations that satisfy i) efficiency, ii) individual rationality, and iii) stability with respect to subcoalitions:

$$C(v) = \{x \in I(v) | \underbrace{\sum_{i \in S} x_i \geq v(S), \forall S \in 2^N \setminus \emptyset}_{\text{stability w.r.t. subcoalitons}} \}.$$

For instance, for a three player game, the core can be described in matrix form by the following set of inequalities:

$$\begin{aligned} C(v) \quad = \quad & \Big\{ x \in \mathbb{R}^3 | \\ & x^T \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & -1 \end{bmatrix} \\ & \geq [v(\{1\})\, v(\{2\}) \ldots v(N) \quad -v(N)] \Big\}. \end{aligned} \tag{6.1}$$

The interpretation of the above conditions is that if $x \in C(v)$ then no coalition has an incentive to split off.

As for the imputation set, one is interested in analyzing conditions for which such stable allocations exist. In this sense, necessary and sufficient conditions for nonemptiness of the Core have been proved by Bondareva, 1963, and Shapley, 1967 and are reiterated below.

Theorem 6.1. (Bondareva & Shapley theorem [Bondareva, 1963, Shapley, 1967]) Given a game $\langle N, v \rangle$, the following are equivalent

1. the core $C(v) \neq \emptyset$,
2. $\langle N, v \rangle$ is a *balanced game.*

The proof which we omit for sake of conciseness is based on duality theorem from linear programming theory.

To understand what a balanced game is, let us introduce the characteristic vector $\mathbf{1}_S$ of coalition S

$$(\mathbf{1}_S)_i = \begin{cases} 1 & i \in S \\ 0 & i \notin S. \end{cases} \tag{6.2}$$

Essentially, for each coalition S, the characteristic vector is a vector with as many components as the number of players. Every component is 1 if the corresponding player is in the coalition and 0 otherwise. Note that the characteristic vectors have already been used in (6.1) as they constitute the columns of the constraint matrix except for the last column. The concept of balanced game builds upon the notion of balanced map which we introduce next.

Definition 6.2. A map $\lambda : 2^N \setminus \emptyset \to \mathbb{R}$ is balanced if

$$\sum_{S \subseteq N, S \neq \emptyset} \lambda(S) \mathbf{1}_S = \mathbf{1}_N.$$

We can interpret a balanced map as a rule summarizing the portion of unitary time that a player gives to each coalition S he belongs to. This is illustrated in the simple example below.

Example 6.1. For a three person game $\lambda(S) = 1/2$ if $|S| = 2$ and 0 otherwise is a balanced map, as well as $\lambda(S) = 1$ if $|S| = 1$ and 0 otherwise. In other words every single player can devote half of his time to any of the two two-player coalitions he forms with the other players, or he could spend all his time working in the grand coalition.

With in mind the notion of balanced map as provided above, we are ready to define balanced games.

Definition 6.3. A game is balanced if for every balanced map λ:

$$\sum_{S \subseteq N, S \neq \emptyset} \lambda(S) v(S) \leq v(N).$$

In the example below, the core coincides with the imputation set.

Example 6.2. Given the game $\langle N, v \rangle$ where $v(S) = 0$ for all $S \neq N$, and $V(N) = 1$, then the Core coincides with the simplex,

$$C(v) = \{x \in \mathbb{R}^n | \, x_i \geq 0, \, \forall i \in N, \, \sum_{i \in N} x_i = 1\}.$$

The example below is called the gloves game as it resemble the case where one player possesses a left glove and the two other players possess a right glove, which naturally gives a higher contractual power to the player with the left glove (player 3 in the example). In this example the core includes only one point.

Example 6.3. (gloves game) Given the *gloves game*: $\langle N, v\rangle$ where

$$v(\{1\}) = v(\{2\}) = v(\{3\}) = v(\{1,2\}) = 0$$
$$v(\{1,3\}) = v(\{2,3\}) = v(\{1,2,3\}) = 1,$$

the core is a singleton,

$$\begin{array}{rcl} C(v) & = & \{x \in \mathbb{R}^3_+ \,|\, x_1 + x_3 \geq 1,\, x_2 + x_3 \geq 1,\, \sum_{i\in N} x_i = 1\} \\ & = & (0,0,1). \end{array} \tag{6.3}$$

6.2 Shapley value

One of the most largely adopted allocation rules by economists is the one formalized by Shapley, and named after him. The underlying idea is strikingly simple and this has contributed to its popularity even among non-experts of the field. The Shapley value is also very intuitive and enjoys useful properties. On the contrary it does not always provide stable allocations. We illustrate the concept by using the following simple story based on the idea that coalitions form according to a pre-ordained sequence.

Suppose that the players enter a room sequentially, and denote by $\sigma : N \to N$ the entry ordering. In other words, $\sigma(k)$ is the player who enters as kth, and $\sigma^{-1}(i)$ is the entry number of player i. For instance, $\sigma = (3,2,1)$ means that

$$\sigma(1) = 3, \quad \sigma(2) = 2, \quad \sigma(3) = 1,$$
$$\sigma^{-1}(1) = 3, \quad \sigma^{-1}(2) = 2, \quad \sigma^{-1}(3) = 1.$$

Let us introduce the set of predecessors of i and denote it by

$$P_\sigma(i) = \{k \in N \,|\, \sigma^{-1}(k) < \sigma^{-1}(i)\}.$$

The marginal value of player i is then given by

$$m_i^\sigma(v) = v(P_\sigma(i) \cup \{i\}) - v(P_\sigma(i)).$$

The marginal value is the additional value that player i contributes when joining the coalition. We collect the marginal values into a single vector, which we call marginal vector and denote as

$$m^\sigma(v) = \{m_i^\sigma(v),\, i \in N\}.$$

Note that we have $n!$ marginal values for player i, that is, as many as the number of permutations of n objects.

The above preamble leads us to the following definition of Shapley vector.

Definition 6.4. (**[Shapley, 1953]**) The Shapley value is the average of the marginal vector over all possible permutations, namely

$$\Phi(v) = \frac{1}{n!} \sum_{\sigma} m^{\sigma}(v).$$

The example below illustrates every single step in the computation of the Shapley value.

Example 6.4. Let the game $\langle N, v \rangle$ be given where

$$v(\{1\}) = v(\{2\}) = v(\{3\}) = 0,$$
$$v(\{1,2\}) = 4,\ v(\{1,3\}) = 7,\quad v(\{2,3\}) = 15,\ v(\{1,2,3\}) = 20.$$

The marginal values are listed in Fig. 6.1 together with the corresponding Shapley value which is $\Phi(v) = \frac{1}{6}(21, 45, 54) \in C(v)$. Let us comment the first column related to $m_1^{\sigma}(v)$ and leave the rest to the reader. If we consider the sequence $(1, 2, 3)$ (first row) this means that player 1 is first, then player 2 joins player 1 and eventually player 3 joins and we have the grand coalition. Now, this implies that the marginal value of player 1 is zero as the player 1's value by playing alone, and before player 2 enters in the coalition is zero. The same consideration applies to the sequence $(1, 3, 2)$ (second row). For the sequence $(2, 1, 3)$ we have that player 2 is first, then player 1 joins him in the coalition and eventually player 3 enters the coalition of 1 and 2. In this case the marginal value of player 1 when entering the coalition with player 2 is 4. This is evident if we consider that the value of $\{1, 2\}$ is 4 while the value of $\{2\}$ is 0. For the sequence $(2, 3, 1)$ (fourth row), player one joins the coalition as last, turning the coalition $\{2, 3\}$ which has a value of 15 into a grand coalition whose value is 20, thus his marginal value is 5. The sequence $(3, 1, 2)$ (in the fifth row) has player 1 entering as second and joining player 3. So player 1 turns the coalition of the single player 3 with zero value into the coalition $\{1, 3\}$ with value 7, thus the latter is also his marginal value. Finally, for the sequence $(3,2,1)$ (last row), player 1 joins as last the coalition of $\{2,3\}$, thus turning the value of the coalition from 15 to 20, which yields a marginal value of 5. Columns 2 and 3 listing the marginal values $m_2^\sigma(v)$ and $m_3^\sigma(v)$ of players 2 and 3 respectively can be obtained by repeating the same reasoning for these players.

σ	$m_1^\sigma(v)$	$m_2^\sigma(v)$	$m_3^\sigma(v)$
(1,2,3)	0	4	16
(1,3,2)	0	13	7
(2,1,3)	4	0	16
(2,3,1)	5	0	15
(3,1,2)	7	13	0
(3,2,1)	5	15	0

Figure 6.1: Marginal values.

The Shapley value has become popular as it is simple to compute, intuitive, and verifies several properties such as i) efficiency, and ii) the "dummy" player property [Hart, 1989]. The latter consists in that each dummy player i is given $\Phi(v)_i = v(\{i\})$, which corresponds to the marginal contribution he creates by entering any coalition. Unfortunately, as anticipated in the introductory paragraph of the chapter, the Shapley value does not always belong to the Core of the game. This is not the case for the nucleolus introduced below, which despite its computation is a bit more troublesome, provides allocation in the core, whenever the latter is nonempty.

6.3 Convex games

Before introducing the nucleolus, we wish to discuss a class of games for which the computation of the core is particularly simplified, i.e., the class of convex games.

Convex games satisfy the property

$$v(S \cup T) + v(S \cap T) > v(S) + v(T),\ \forall S, T \in 2^N.$$

Let us recall that for superadditive games we have

$$v(S \cup T) > v(S) + v(T),\ \forall S, T \in 2^N : S \cap T = \emptyset,$$

then convex games are a subset of superadditive games. Essentially, in convex games the marginal contribution of any fixed player i or set T to coalition S rises as more players join S. Given a convex game, any allocation x such that $x_i = m_i^\sigma$, whatever σ, yields a point in the Core. We are ready to illustrate the nucleolus in the remaining part of this chapter.

6.4 Nucleolus

Let us introduce the lexicographic order (exactly the one used to store words in a dictionary) $\leq_L$ where $x \in \mathbb{R}^p$ is lexicographically smaller than $y \in \mathbb{R}^P$ if $x = y$ or there exists an $s = 1, 2, \dots, p$ such that $x_i = y_i$ for all $i < s$ and $x_s < y_s$. For instance, we can say that $(0, 100, 100) \leq_L (1, -10, -10)$, and $(10, 4, 100) \leq_L (10, 5, 6)$.

Let us define the excess vector of a coalition S for given $x \in I(v)$:

$$\theta_x = \{e(S, x) := v(S) - \sum_{i \in S} x_i,\ \forall S \in 2^N \setminus \emptyset\}.$$

In plain words, the excess vector has as many components as the number of coalitions, and for each coalition it stores the deviation between the value of the coalition and the total amount given to the members of the coalition. For the coalition to be stable we expect such a difference to be negative, in which case the total amount exceeds the value of the coalition. Given the lexicographic ordering defined above, the nucleolus is the solution that minimizes such excess vector as formalized next.

Definition 6.5. The Nucleolus is the *lexicographic minimizer* of any complaint vector (decreasing order):

$$\theta(Nu(v)) \leq_L \theta(x),\ \forall x \in I(v).$$

A seminal result which we report below establishes that the nucleolus is always in the core of the game whenever the latter is nonempty.

Theorem 6.2. If $C(v) \neq \emptyset$, then $Nu(v) \in C(v)$.

In other words, the nucleolus minimizes the maximal complaint and if the core is nonempty, the nucleolus always belongs to it. For the gloves game introduced the nucleolus then corresponds to the core, which we remind is a singleton.

Example 6.5. Given the *gloves game*: $\langle N, v \rangle$ where

$$\begin{aligned} v(\{1\}) = v(\{2\}) = v(\{3\}) = v(\{1,2\}) = 0 \\ v(\{1,3\}) = v(\{2,3\}) = v(\{1,2,3\}) = 1, \end{aligned}$$

the nucleolus is $Nu(v) = (0, 0, 1) = C(v)$.

If the game admits a symmetric structure the computation of the nucleolus can be enormously simplified as in the following example.

Example 6.6. Given the game $\langle N, v \rangle$, $N = \{1, 2, 3, 4\}$ and $v(S) = |S|^2$ for all $S \in 2^N$,

$$\begin{aligned} C(v) \quad = \quad & \{x \in \mathbb{R}^4_+ |\, x_i + x_j \geq 4,\, \forall i, j,\, x_i + x_j + x_k \geq 9,\, \forall i, j, k, \\ & x_1 + x_2 + x_3 + x_4 = 16\}. \end{aligned} \tag{6.4}$$

Note that the constraints are symmetric thus $Nu(v) = (a, a, a, a)$. The only solution verifying efficiency is then $Nu(v) = (4, 4, 4, 4)$.

6.4.1 Computation through sequence of linear programs

In this last part of the chapter we wish to describe a computational technique to calculate the nucleolus. In particular, the nucleolus can be computed through a sequence of linear programming problems. Let us illustrate the technique in a recursive way by looking at the first two steps. Step 1 involves minimizing the maximal excess, and thus we need to solve the following linear programming problem:

$$\begin{aligned} & \theta_1 := \min t \\ & e(S, x) \leq t,\, \forall S \subseteq N, \\ & \mathbf{1}^T x(N) = v(N), \\ & x_i \geq v(\{i\}),\, \forall i \in N. \end{aligned} \tag{6.5}$$

Denote by $X_1 = \{x | e(S, x) \leq \theta_1, \forall S \subseteq N\}$, i.e., X_1 is the set of allocations that attain the minimum for the maximal excess. Also, denote by $\Sigma_1 = \{S \subseteq N | e(S, x) = \theta_1, \forall x \in X_1\}$. Σ_1 is the set of all coalitions at which the maximal excess is attained at all $x \in X_1$. We can then move to the next step. Step 2 consists in minimizing the second-largest excess, which boils down to the following linear programming problem:

$$\begin{aligned} & \theta_2 := \min t \\ & e(S, x) = \theta_1, \forall S \in \Sigma_1, \\ & e(S, x) \leq t, \forall S \notin \Sigma_1, \\ & \mathbf{1}^T x(N) = v(N), \\ & x_i \geq v(\{i\}), \forall i \in N, \quad \text{and so forth.} \end{aligned} \tag{6.6}$$

We can then proceed recursively until a solution is obtained, and such a solution is the nucleolus of the game. One can note that, despite the advantages shown by the nucleolus (if compared with the Shapley value) which is always in the core of the game, the computation of the nucleolus is much more complicated than the one of the Shapley value which motivates the large popularity of the second one among the non-experts of game theory.

7

Evolutionary game theory

This chapter, following the open course by Benjamin Polak at Yale [Polak, 2007], covers basic concepts in the theory of evolutionary games. We shall introduce the fundamental concept of *evolutionary stable strategies (ESS)*, and for it we provide two equivalent formal definitions; the first one is borrowed from evolutionary biology while the second one is developed in economics and is linked back to the NE. ESS are illustrated on classical examples like the Prisoner's dilemma, the Hawk and Dove game, the Coordination game, just to name a few.

7.1 Historical note

The foundations of evolutionary game theory were laid by John Maynard Smith in his paper "Game Theory and the Evolution of Fighting" [Smith, 1972], and one year later in the article "The Logic of Animal Conflict", coauthored by G. Price [Smith and Price, 1973], and finally in the book "Evolution and the Theory of Games" [Smith, 1982].

The theory of evolutionary games studies the influence of game theory on evolutionary biology, as well as the influence of evolutionary biology on social science as illustrated next.

7.2 Model and preliminary considerations

Consider a large population of individuals, called *incumbents*, who are "programmed" to play a given strategy. At every time we have random matchings between individuals, which result in a 2-player symmetric game where every individual tries to maximize its average payoff. Strategies are *genes*, and payoffs indicate the *fitness* of the individuals, namely, the expected number of offsprings. Mutations in the population consist in some offsprings, the *mutants*, playing randomly over the set of feasible strategies. Successful strategies tend to grow while unsuccessful strategies tend to extinguish.

In this context, a main question is the following one: Is a given strategy robust against mutations? Are mutants going to die or thrive?

Example 7.1. (Prisoner's Dilemma) Let us illustrate the evolutionary game theoretic scenario on the Prisoner's Dilemma in Fig. 7.1. Suppose we have a group of lions on a hunt, or ants defending a nest. Every lion can choose to "cooperate" and go on a hunt together with the group or "defect" and go on a hunt alone. Likewise, every ant can cooperate in defending the nest from a spider or can defect and flee away from the danger.

	C	*D*
C	(3,3)	(0,4)
D	(4,0)	(1,1)

Figure 7.1: Prisoners' dilemma as evolutionary game.

The bimatrix in Fig. 7.1 depicts the payoff of every single individual, say P_1 when he plays randomly against an individual from the population. Given this, consider a mixed strategy $(1-\epsilon, \epsilon)$ for P_2 and interpret this as if a small portion ϵ of mutants play D while the rest of the population (incumbents) play C. We wish to answer the question whether "cooperation" is *evolutionarily stable* (ES). The answer is negative, that is, C is not ES as the mutant performs better than the incumbent on a random matching. To see this, observe that when an incumbent (row player) meets an opponent (column player) randomly extracted from the population, the opponent player will cooperate (pick C) with probability $1-\epsilon$ and will defect (pick D) with probability ϵ. In other words it is like if the opponent player would adopt a mixed strategy $(1-\epsilon, \epsilon)$. The expected payoff is then $(1-\epsilon)3+\epsilon 0 = 3(1-\epsilon)$. In a nutshell we have:

$$C \quad \text{vs.} \quad [(1-\epsilon)C + \epsilon D] \to (1-\epsilon)[3] + \epsilon 0 = 3(1-\epsilon).$$

Differently, when a mutant (row player) meets an opponent (column player) randomly extracted from the population, again the opponent player will cooperate (pick C) with probability $1-\epsilon$ and will defect (pick D) with probability ϵ and the expected payoff is then $(1-\epsilon)4+\epsilon = 4(1-\epsilon)+\epsilon$. This is summarized in the line below:

$$D \quad \text{vs.} \quad [(1-\epsilon)C + \epsilon D] \to (1-\epsilon)[4] + \epsilon 1 = 4(1-\epsilon) + \epsilon.$$

Now, comparing the expected payoffs of incumbents and mutants when involved on random matchings with other individuals, we see that the mutant performs better than the incumbent, from which we conclude that "cooperation" is not an ESS. Note that "cooperation" is a strictly dominated strategy so the question is now whether we can infer some general result from the given example. In the next section we try to answer this question.

7.2.1 Strictly dominated strategies are not ES

In the Prisoner's dilemma example discussed in the preceding section, we saw that cooperation, which is a strictly dominated strategy, is not ES. Here we generalize the idea and illustrate a main fact according to which *strictly dominated strategies are not ES*. To see this, consider again the Prisoner's Dilemma example, and assume that the population is programmed to defect, while there is a minority who cooperates and picks C. This corresponds to assuming a mixed strategy $(\epsilon, 1-\epsilon)$ for P_2 (column player) and we can interpret this as if a small portion ϵ of mutants play C while incumbents play D. We see here that the strategy "defection" is ES, as a mutant performs worse than an incumbent on a random matching. This is illustrated next: assume that an incumbent (row player) randomly matches with another individual (column player) extracted from the population, then its expected payoff is $(1-\epsilon)[1]+\epsilon[4]=(1-\epsilon)+4\epsilon$, namely:

$$D \quad \text{vs.} \quad [(1-\epsilon)D+\epsilon C] \rightarrow (1-\epsilon)[1]+\epsilon[4]=(1-\epsilon)+4\epsilon.$$

Likewise, suppose that a mutant (row player) plays against an opponent randomly extracted from the population; his expected payoff is then $(1-\epsilon)[0]+\epsilon[3]=3\epsilon$. More compactly we have:

$$C \quad \text{vs.} \quad [(1-\epsilon)D+\epsilon C] \rightarrow (1-\epsilon)[0]+\epsilon[3]=3\epsilon.$$

We observe that the incumbent is more successful than the mutant on random matchings. In other words any mutation from D extinguishes. Then, the conclusion is that a strictly dominated strategy is not ES as the strictly dominant strategy will be a successful mutation.

7.2.2 ES strategies yield symmetric NE solutions

A second general consideration we wish to point out here is that ES strategies yield symmetric NE solutions. Recalling that strategies are genes, a symmetric NE corresponds to a *monomorphic* population, i.e., a population with a unique gene. This concept is illustrated in the bimatrix example below

	b	c
b	(0,0)	(1,1)
c	(1,1)	(0,0)

The question we wish to answer is whether the strategy c is ES or not. The answer is negative, as we can see that mutants who play b perform better than the incumbents who play c, as it is summarized in the next two lines:

$$\begin{aligned} c \quad \text{vs.} \quad [(1-\epsilon)c+\epsilon b] &\rightarrow (1-\epsilon)[0]+\epsilon[1]=\epsilon, \\ b \quad \text{vs.} \quad [(1-\epsilon)c+\epsilon b] &\rightarrow (1-\epsilon)[1]+\epsilon[0]=1-\epsilon. \end{aligned}$$

In the first line above, an incumbent playing c matches with an opponent playing the mixed strategy $[(1-\epsilon)c+\epsilon b]$ and the expected payoff is then $(1-\epsilon)[0]+\epsilon[1]=\epsilon$. The second line refers to the case where a mutant playing b matches with an opponent from the population who plays the mixed strategy $[(1-\epsilon)c+\epsilon b]$ and the expected payoff is then $(1-\epsilon)[1]+\epsilon[0]=1-\epsilon$. Then, the mutant's payoff is higher than the incumbent's payoff from which we conclude the thesis.

After careful consideration we also note that mutants playing b, who are more successful than incumbents playing c, will grow from a small proportion ϵ to $1/2$. Also note that b, the mutant or invader, is itself not ES, though it still avoids dying out. The question is now whether is c an NE or not. In other words is (c,c) a symmetric NE? Once more the answer is negative since b is a strict profitable deviation. So we can conclude that if a generic strategy s does not yield a symmetric NE, then s is not ES. In other words, being an NE is a necessary condition for a strategy to be ES. In a nutshell we then have the following implication:

$$\text{If } s \text{ is ES} \quad \Rightarrow \quad (s,s) \text{ is NE.}$$

7.2.3 An NE strategy is not necessarily ES

In the previous section we saw that an ES strategy yields a symmetric NE. Here, we show that the converse is not true, i.e., an NE strategy is not necessarily ES. Consider the example below:

	a	b
a	(1,1)	(0,0)
b	(0,0)	(0,0)

We have two NE solutions, (a,a) and (b,b). However the strategy b is not ES. To see this note that mutants playing a perform better than incumbents playing b:

$$\begin{aligned} b \quad &\text{vs.} \quad [(1-\epsilon)b+\epsilon a] \to (1-\epsilon)[0]+\epsilon[0]=0 \\ a \quad &\text{vs.} \quad [(1-\epsilon)b+\epsilon a] \to (1-\epsilon)[0]+\epsilon[1]=\epsilon. \end{aligned}$$

In the first line above, an incumbent b matches with an opponent from the population playing the mixed strategy $[(1-\epsilon)b+\epsilon a]$ and the resulting expected payoff is $(1-\epsilon)[0]+\epsilon[0]=0$. The second line refers to the case where a mutant playing a matches with an opponent from the population playing $[(1-\epsilon)b+\epsilon a]$ and the resulting expected payoff is then $(1-\epsilon)[0]+\epsilon[1]=\epsilon$. Then, the mutant beats the incumbent and we can conclude that though (b,b) is an NE the strategy b is not ES. Intuitively, the reason behind this is that (b,b) is not a strict NE. Actually we can observe the following fact: If (s,s) is a strict NE then s is ES!

7.3 Formal definition of ESS

In this section we provide two formal and equivalent definitions of ESS, one from evolutionary biology, the second from economics. While the first one builds upon the notion of "small" perturbation ϵ, which makes the verification of the conditions a bit troublesome, the definition from the economics is more direct and easier to be verified.

7.3.1 Formal definition from biology

Consider a 2-player symmetric game where Δ is the set of mixed strategies for a single player, the row player, and $u(a,b)$ is its payoff when it plays the strategy $a\in\Delta$ against a population playing $b\in\Delta$.

Definition 7.1. A mixed strategy $s^*\in\Delta$ is ES if there exists $\bar{\epsilon}>0$ such that for any $s\in\Delta$ and $\epsilon\leq\bar{\epsilon}$, we have

$$\underbrace{u(s^*,\epsilon s+(1-\epsilon)s^*)}_{\text{payoff to ES } s^*} > \underbrace{u(s,\epsilon s+(1-\epsilon)s^*)}_{\text{payoff to mutant } s}.$$

A first interpretation is that the incumbents perform better than the mutants on random matchings. A second interpretation is that strategy s^* cannot be invaded by s.

7.3.2 Formal definition from economics

Economists have come across the same concept of ESS and provide an equivalent definition which reminds the first and second condition of optimality.

Definition 7.2. A mixed strategy $s^*\in\Delta$ is ES if for any $s\in\Delta$ we have

(a) $u(s^*, s^*) \geq u(s, s^*)$ ((s^*, s^*) is symmetric NE), and

(b) if $u(s^*, s^*) = u(s, s^*)$ ((s^*, s^*) not strict NE), then $u(s^*, s) > u(s, s)$ (mutant performs bad against itself).

The above means that (a) the mutant performs poorly against the masses, and also that (b) the mutant performs reasonably well against the masses but badly against itself. It can be proven that the two definitions are equivalent.

7.3.3 From (b), a non strict NE can be ES

The second definition has an immediate consequence: A non strict NE can be ES. This can be shown in the example below.

	a	b
a	(1,1)	(1,1)
b	(1,1)	(0,0)

The pair (a, a) is a symmetric NE but not a strict NE as $u(a, a) = u(b, a) = 1$. Then we pose the question: Is a ES? To give an answer, let us check condition (b):

$$u(a, b) > u(b, b),$$

from which we conclude that a is ES.

Example 7.2. (Evolution of social convention) We can have multiple ES strategies. The coordination game below is commonly used to describe the evolution of social conventions.

	L	R
L	$(2, 2)$	$(0, 0)$
R	$(0, 0)$	$(1, 1)$

A classical example is the "drive" convention, namely whether driving left or right (note that this is exactly a coordination game). One notes two strict NE solutions, (L, L) and (R, R). One also observes that both L and R are ES. Then we conclude that we can have multiple ES strategies. These need not be equally good, for instance in the example (L, L) is better than (R, R).

Example 7.3. (Battle of the sexes) The current example shows that we can have ES mixed strategies. Recalling that strategies are genes, in evolutionary biology language this corresponds to a *polymorphic* population, i.e., a population with multiple genes. This concept is illustrated in the bimatrix below which is a slight variation of the Battle of the sexes example obtained from swapping the columns.

	a	b
a	(0,0)	(2,1)
b	(1,2)	(0,0)

A possible interpretation for the example above is that we have two possible strategies: a is aggression, b is non aggression. The game admits no symmetric pure NE (monomorphic population).

However, the solution $[(2/3, 1/3), (2/3, 1/3)]$ is a symmetric mixed-strategy NE (polymorphic population). It cannot be strict as it is a mixed NE (any strategy in the support returns same payoff). From (b), we can check that $u(s^*, s) > u(s, s)$ for all possible mutations $s \in \Delta$. Indeed, take $s = a$, and $s = b$ and obtain for the row player

$$u(s^*, a) = 1/3 > u(a, a) = 0, \quad u(s^*, b) = 4/3 > u(b, b) = 0.$$

Example 7.4. The Hawk and Dove game: monomorphic ESS (case 1) This example shows that we can have monomorphic ESS.

	$Hawk$	$Dove$
$Hawk$	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
$Dove$	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

In the example V is the prize of victory, and C is the cost of fight. If $V > C$, then there is a unique strict Nash equilibrium (like in the Prisoner's dilemma), which is $(Hawk, Hawk)$. Therefore, $Hawk$ is also ES. If the prize of victory is higher than the cost of fighting then everybody will end up fighting.

Example 7.5. The Hawk and Dove game: monomorphic ESS (case 2) In this example we show that Hawk is a monomorphic ESS.

	$Hawk$	$Dove$
$Hawk$	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
$Dove$	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

Actually, if $V = C$, then $(Hawk, Hawk)$ is still an NE but not strict as $u(D,H) = u(H,H) = 0$. Is $Hawk$ ES? Yes, as for any mutation D, $u(H,D) = V > u(D,D) = V/2$. If the prize of victory is equal to the cost of fighting then everybody will end up fighting.

Example 7.6. The Hawk and Dove: polymorphic ESS (case 3) This example shows that we can have polymorphic ESS.

	Hawk	Dove
Hawk	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	(V,0)
Dove	(0,V)	$\left(\frac{V}{2}, \frac{V}{2}\right)$

Actually, if $V < C$, (H,D) and (D,H) are two non symmetric NE solutions, and we have a mixed NE $[(V/C, 1 - V/C), (V/C, 1 - V/C)]$. One can note that $Hawk$ is not ES, and $Dove$ is not ES. Is $s^* = (V/C, 1 - V/C)$ ES? As $[(V/C, 1 - V/C), (V/C, 1 - V/C)]$ is a mixed NE then it is not strict, $u(D, s^*) = u(H, s^*) = u(s^*, s^*)$ Then we can check first that $u(s^*, H) > u(H,H)$ (for a mutant H), and second that $u(s^*, D) > u(D,D)$ (for a mutant D). Note that as V increases then more Hawks are in the ESS.

Example 7.7. Rock-paper-scissors game: no ESS. This last example points out that we can have no ESS.

	R	P	S
R	(γ, γ)	$(-1, 1)$	$(1, -1)$
P	$(1, -1)$	(γ, γ)	$(-1, 1)$
S	$(-1, 1)$	$(1, -1)$	(γ, γ)

This is another example where NE does not imply ESS. The game admits one mixed (therefore non strict) NE at $s^* = (1/3, 1/3, 1/3)$. Indeed, we recall that from the indifference principle it must hold

$$u(R, s^*) = u(P, s^*) = u(S, s^*) = u(s^*, s^*).$$

It is immediate to verify that the mutant R is successful, i.e.,

$$u(s^*, R) = \gamma/3 < u(R,R) = \gamma.$$

We can then conclude that the game admits no ESS. Actually the strategies keep cycling around. This example is abundantly used in the theory of learning in games and uncoupled dynamics to prove that convergence through learning is not always possible which is the subject of the next chapter.

8

Replicator dynamics and learning

So far we have discussed different types of equilibria, which from a system theoretic perspective represent stationary solutions, without addressing explicitly the kind of dynamics that can lead to such equilibria. With this chapter we introduce for the first time a dynamics, known as *replicator dynamics* [Sandholm, 2010, Weibull, 1995], describing the time evolution of the strategies when the players are keen to react to what they observe using their best response strategies. The topic offers the opportunity to elaborate on asymptotically stable solutions and their connection with the ESS defined in the previous chapter. The second part of this chapter deals with learning processes in game theoretic contexts [Fudenberg and Levine, 1998]. The treatment of the topic follows the open MIT course on Game Theory with Engineering Applications by Asuman Ozdaglar [Ozdaglar, 2010]. Learning in games is another field where dynamical aspects play a major role. In particular, we shall discuss situations where the players observe the past opponents' play and infer a probability distribution, called *empirical frequency*, over the opponents' action spaces. Such a scenario is known in the literature as *fictitious game.*

8.1 Replicator dynamics: set-up

Our starting observation is that there is no explicit dynamics in the evolutionary models seen so far. To introduce a dynamics, let us enumerate strategies by $s = 1, 2, \dots, K$, and denote the fraction of the population playing s by x_s, where

$$\sum_{s=1}^{K} x_s = 1.$$

Also, denote $x = (x_s)_{s \in 1,2,\dots,K}$, the population distribution, or polymorphic strategy profile. This can be viewed as a mixed strategy over $1, 2, \dots, K$. Assume random matchings between individuals in the population. As in the theory of evolutionary games covered in the previous chapter, what really matters is the expected fitness of playing s against a population playing x, which is given by $u(s, x)$.

The replicator equation in its generic form is then given by, for each $s = 1, 2, \dots, K$ and for all t and τ

$$x_s(t+\tau) - x_s(t) = x_s(t) \frac{\tau[u(s, x(t)) - \bar{u}(x(t))]}{\bar{u}(x(t))},$$

where $\bar{u}(x(t))$ is the average fitness at time t when the population play $x(t)$, and is obtained as

$$\bar{u}(x(t)) := \sum_{s=1}^{K} x_s u(s, x(t)).$$

We can interpret the above equation as follows: the greater the fitness of a strategy relative to the average fitness, the greater its relative increase in the population. Note that $\sum_{s=1}^{K} x_s(t+\tau) = 1$. Also note that if we take $\tau = 1$ we have the well known *discrete-time replicator equation*.

The *replicator equation in continuous-time* can be derived by dividing both sides of the replicator equation by τ and taking the limit as $\tau \to 0$. This gives

$$\lim_{\tau \to 0} \frac{x_s(t+\tau) - x_s(t)}{\tau} = x_s(t) \frac{[u(s, x(t)) - \bar{u}(x(t))]}{\bar{u}(x(t))}.$$

This yields what is known as the *continuous replicator*

$$\dot{x}_s(t) = x_s(t)\frac{[u(s, x(t)) - \bar{u}(x(t))]}{\bar{u}(x(t))}.$$

The given context raises several questions which we will address later on: Is a given vector of distribution x^* a stationary state, i.e., $\dot{x}^*(t) = 0$? Is a given vector of distribution x^* an asymptotically stable state? Namely, does there exist a neighborhood of x^* such that starting from any x_0 in this neighborhood, the continuous replicator dynamics converges to x^*? Let us start by addressing stationarity for the continuous replicator dynamics introduced above.

8.1.1 An NE is a stationary state

We start by observing that an NE is a stationary state for the replicator dynamics. This is formalized in the following theorem.

Theorem 8.1. If x^* is an NE, then it is a stationary state.

The proof is based on the following idea. If x^* is an NE, then it is a best response to itself, and thus

$$\begin{aligned} &u(s, x(t)) - \bar{u}(x(t)) \leq 0 && \text{for all } s \\ &u(s, x(t)) - \bar{u}(x(t)) = 0 && \text{for all } s \text{ in the support of } x^*. \end{aligned}$$

Thus for any s, either $u(s, x(t)) - \bar{u}(x(t)) = 0$ or $x_s(t) = 0$, and hence $\dot{x}_s(t) = 0$ for all s. Note that the converse is not true: actually, let x^* be a non-Nash pure strategy, then $x_s(t) = 0$ for all s other than the pure strategy in question, and x^* is stable.

8.1.2 Asymptotic stability implies NE

We next discuss the fact that asymptotic stability implies NE.

Theorem 8.2. If x^* is asymptotically stable, then it is an NE.

The proof is immediate if x^* corresponds to a pure strategy (monomorphic population). In the case where x^* corresponds to a mixed strategy Nash equilibrium, the proof is also straightforward but long. The basic idea is that the continuous replicator equation implies that we are moving in the direction of better responses relative to the average. If this process converges, then there must not exist any more (any other) strict better responses, and thus we must be at a Nash equilibrium. The converse is again not true.

To see that the converse is not true consider the following example.

Example 8.1. Let the following bimatrix game be given

	a	b
a	(1,1)	(0,0)
b	(0,0)	(0,0)

The pair (b, b) is an NE, but such an equilibrium is not asymptotically stable. Indeed, any perturbation away from (b, b) will start a process in which the fraction of agents playing a steadily increases. To see this, take $x(t) = (\epsilon, 1 - \epsilon)$, then

$$x_a = \epsilon, \quad u(a, x(t)) - \bar{u}(x(t)) = \epsilon - \epsilon^2 > 0.$$

From the above we understand that playing a is more successful in comparison with the average payoff computed over the population and therefore the percentage of people playing a increases.

8.1.3 ES implies asymptotic stability

The next result shows that there is a one direction implication between evolutionary stability and asymptotical stability.

Theorem 8.3. If x^* is ES, then it is asymptotically stable.

Sketch of Proof. Recall the definition of ESS which we rewrite below.

A mixed strategy $s^* \in \Delta$ is ES if there exists $\bar{\epsilon} > 0$ such that for any $s \in \Delta$ and $\epsilon \leq \bar{\epsilon}$, we have

$$\underbrace{u(s^*, \epsilon s + (1-\epsilon)s^*)}_{\text{payoff to ES } s^*} > \underbrace{u(s, \epsilon s + (1-\epsilon)s^*)}_{\text{payoff to mutant } s}.$$

Let A be the game matrix, namely the matrix involving the payoffs of the row player (we recall here that in this chapter we consider only symmetric games if not differently specified). The idea is then to rewrite the above condition as

$$x^{*T} Ax > x^T Ax, \quad \forall x \text{ in neighborhood of } x^*.$$

The above inequality yields

$$\underbrace{(x^* - x)}_{z} A(x - x^*) > 0$$

or

$$V(z) = zAz < 0 \text{ (negative definite game)}$$

where $\sum_i z_i = 0$. It remains to observe that $\dot{V}(z) = zA\dot{z} > 0 \ \forall z \neq 0$ when $\dot{z}$ is obtained from replicator dynamics.

8.2 Learning in games

A second field where dynamical aspects come into play in a predominant way is learning in games. In a learning context we assume that the game is played repeatedly. Individuals respond to what they observe and hopefully their play converges to a Nash equilibrium. We suppose that the players do not know the opponents' payoffs. This leads to the so-called *uncoupled dynamics*: "my updates do not depend on others' payoffs" [Hart, 2005, Hart and Mas-Colell, 2001, 2003]. Note that in replicator dynamics one needs to know the average payoff. The example below simulates a typical learning scenario.

Example 8.2. (How would you play?) Suppose the row player has to select an action in the game below. He knows his payoffs but ignores his opponent's payoffs.

	Stag	*Hare*
Stag	$(10, \times)$	$(0, \times)$
Hare	$(2, \times)$	$(3, \times)$

Now, suppose that the game has already been played 10 times; the column player has selected *Stag* 7 times and *Hare* 3 times. Based on this experience, the column player appears to be playing a mixed strategy $z = (0.7, 0.3)$. The best response to z is then

$$\begin{array}{rcccl} u_1(Stag, z) & = & 0.7 \cdot 10 + 0.3 \cdot 0 & = & 7 \\ u_1(Hare, z) & = & 0.7 \cdot 2 + 0.3 \cdot 3 & = & 2.3. \end{array} \tag{8.1}$$

If the column player will continue as before, the row player is best to select *Stag*.

8.2.1 Fictitious play

A relevant branch of the theory of learning in games is represented by the well known *fictitious play*. Here, we assume that the players are *myopic*, that is they always play a best response to their best guess of the opponent's mixed strategy. Now the question arises of what this *best guess* is. To answer this question we make an assumption playing around the concept of *stationarity*, namely, every player thinks that his opponent's actions have all been selected according to a fixed mixed strategy. In other words, the "opponent's strategy is stationary".

Given such a preamble, in order to develop a mathematical model for the learning process, let us consider the tuple $\langle N, (\mathcal{A}_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where N is the set of players, and $\mathcal{A}_i$ is the set of pure actions of player i. The set of action profiles is $A := \{a = (a_i)_{i \in N},\, a_i \in \mathcal{A}_i(s)\}$. Players play repeatedly at time $t = 1, 2, \ldots$. The payoff function of player i at each time is $u_i : A \to \mathbb{R}$. For each $a_{-i} \in A_{-i}$, let $\kappa_i^t(a_{-i})$ be the number of plays on which individual i has observed his opponent playing a_{-i}. Also, let $\kappa_i^0(a_{-i})$ be starting point, fictitious past.

Example 8.3. Consider a two-player game, with $\mathcal{A}_2 = \{U, D\}$. If $\kappa_1^0(U) = 3$ and $\kappa_1^0(D) = 5$, and player 2 plays U, U, D in the first three periods, then $\kappa_1^3(U) = 5$ and $\kappa_1^3(D) = 6$.

The main idea in fictitious play is that each player assumes that his opponent is using a stationary mixed strategy, and updates his beliefs about this stationary mixed strategies at each step. Players choose actions in each period (or stage) to maximize the period's expected payoff given their prediction of the distribution of their opponent's actions, which they form according to:

$$\mu_i^t(a_{-i}) = \frac{\kappa_i^t(a_{-i})}{\sum_{a_{-i} \in \mathcal{A}_{-i}} \kappa_i^t(a_{-i})}.$$

For example, in a two-player game, player i forecasts player $-i$'s strategy at time t to be the *empirical frequency* distribution of past play. Player i selects the best response to $\mu_i^t(a_{-i})$, which is given by:

$$a_i^t = \arg\max_{a_i \in \mathcal{A}_i} u_i(a_i, \mu_i^t(a_{-i})) \in \mathcal{B}_i(\mu_i^t(a_{-i})).$$

We can then derive the *beliefs' iteration* as follows:

$$\begin{aligned} \mu_i^t(a_{-i}) &= \frac{\kappa_i^t(a_{-i})}{t} = \frac{\kappa_i^{t-1}(a_{-i}) + \mathbf{1}_{\{a_{-i}^t = a_{-i}\}}}{t} \\ &= \left(\frac{t-1}{t}\right)\frac{\kappa_i^{t-1}(a_{-i})}{t-1} + \frac{\mathbf{1}_{\{a_{-i}^t = a_{-i}\}}}{t} \\ &= \left(\frac{t-1}{t}\right)\mu_i^{t-1}(a_{-i}) + \frac{\mathbf{1}_{\{a_{-i}^t = a_{-i}\}}}{t}. \end{aligned} \tag{8.2}$$

Observe that there is no need to store all past plays. Also, note that we can perform online estimation in that the current belief depends on previous belief and current observation. The following example shows a few steps of a learning process in fictitious play.

Example 8.4. Consider the fictitious play of the following game:

	L	R
U	$(3,3)$	(0,0)
D	(4,0)	$(1,1)$

Note that this game is dominant solvable (D is a strictly dominant strategy for the row player), and the unique NE is (D, R). Assume that $\kappa_1^0 = (3, 0)$ and $\kappa_2^0 = (1, 2.5)$. Then the fictitious play proceeds as follows:

- Period 1: Then, $\mu_1^0 = (1, 0)$ and $\mu_2^0 = (1/3.5, 2.5/3.5)$, so play follows $a_1^0 = D$ and $a_2^0 = L$.
- Period 2: we have $\kappa_1^1 = (4, 0)$ and $\kappa_2^1 = (1, 3.5)$, so play follows $a_1^1 = D$ and $a_2^1 = R$.
- Period 3: we have $\kappa_1^2 = (4, 1)$ and $\kappa_2^2 = (1, 4.5)$, so play follows $a_1^2 = D$ and $a_2^2 = R$ and so forth.

From the above example we understand that convergence may occur. Actually, since D is a dominant strategy for the row player, he always plays D, and μ_2^t converges to $(0,1)$ with probability 1. Therefore, player 2 will end up playing R. The remarkable feature of the fictitious play is that players do not have to know anything about their opponent's payoff. They only form beliefs about how their opponents will play.

Generalizing our discussion, convergence is taken to occur if $\mu^t = (\mu_i^t)_{i \in N} \to \mu^*$. We also note that if convergence occurs then μ^* must be a Nash equilibrium. Games in which convergence must occur are said to have the fictitious play property. However, in general convergence does not always occur: see [Shapley, 1964] and [Jordan, 1993].

9

Differential games

This chapter follows the tutorial written by Alberto Bressan on "Non-cooperative Differential Games" in 2010 [Bressan, 2010]. Differential games are games with an underlying state variable evolving according to a differential equation taking the players' action as input. The payoff of the players depend not only on the players action, but also on the current value of the state. As differential games are a generalization of optimal control problems, we first review the foundations of optimal control theory, and in particular, the Pontryagin Maximum Principle (PMP) and the Hamilton-Jacobi-Bellman (HJB) equation. We then shall consider differential games under open-loop or closed-loop strategies before concluding by introducing the Hamilton-Jacobi-Isaacs equation. The last part of the chapter covers linear quadratic differential games and sheds light on connections with H^∞-optimal control.

9.1 Optimal control problem

Let a controlled dynamics be given as below, where f is continuous w.r.t (x, u, t) and differentiable w.r.t. x, $x(t) \in \mathbb{R}^m$ is the state, $u(t) \in U$ is

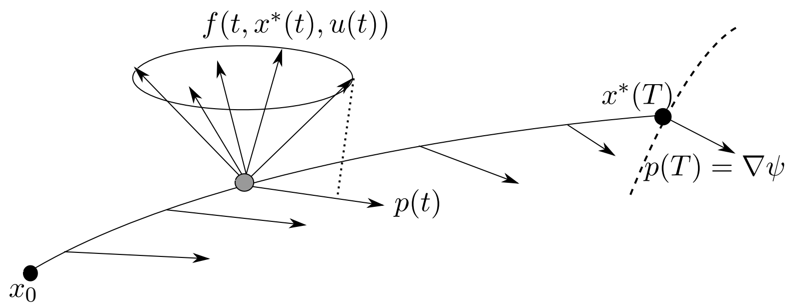

Figure 9.1: Graphical illustration of the PMP when the running cost is null. Courtesy by Alberto Bressan, Noncooperative Differential Games. A Tutorial (2010).

the control, $U \subseteq \mathbb{R}^m$ is compact:

$$\dot{x}(t) = f(t, x(t), u(t)), \quad u(t) \in U.$$

Note that compactness is necessary to guarantee the existence of an optimal solution, see [Liberzon, 2012, Chap. 4.5].

Assume that the solutions in finite time are not unbounded, i.e.,

$$|f(t, x, u)| \leq C(1 + |x|), \quad \forall\, (t, x, u) \in [0, T] \times \mathbb{R}^m \times U.$$

Now, given $x(t_0) = x_0$, denote the trajectory starting from x_0 at t_0 by

$$t \mapsto x(t) = (t; t_0, x_0, u).$$

Consider the optimization problem:

$$\max\ J(u; t_0, x_0) := \psi(x(T)) - \int_{t_0}^{T} L(t, x(t), u(t))dt,$$

where the "max" operation is over measurable functions $u : [t_0, T] \mapsto U$.

9.1.1 Pontryagin Maximum Principle (PMP)

A first seminal result borrowed from the Russian literature of the 50s is the world well known PMP which we illustrate in the following. Note that the PMP is a "necessary" condition for a control to be optimal.

Consider a free terminal point problem, namely, a problem where the final state is not constrained to take any specific value. Let $u^*(t)$ be the optimal control, $x^*(t)$ be the corresponding trajectory and $p(t)$

the adjoint variable satisfying:

$$\dot{p}(t) = -p(t)\frac{\partial f}{\partial x}(t, x^*(t), u^*(t)) + \frac{\partial L}{\partial x}(t, x^*(t), u^*(t)),\ p(T) = \nabla\psi(x^*(T)).$$

The following maximality condition holds

$$u^*(t) = \arg \underbrace{\max_{u \in U} \Big\{ p(t) \cdot f(t, x^*(t), u(t)) - L(t, x^*(t), u(t)) \Big\}}_{H(t,x,p) \text{ is the maximized Hamiltonian}},$$

where ∇ denotes the gradient operator. A graphical illustration of the PMP is given in Fig. 9.1 for the case where the running cost is null. Essentially, the maximality condition returns at every time the control corresponding to the maximal inner product between the RHS of the dynamics and the adjoint variable.

9.1.2 Two-point boundary value problem

The PMP is also "sufficient" if the *maximized Hamiltonian* is concave on x, where the maximized Hamiltonian is given by

$$H(t, x, p) := \max_{u \in U} \Big\{ p(t) \cdot f(t, x(t), u(t)) - L(t, x(t), u(t)) \Big\}.$$

Then one can rearrange the problem as follows. First we solve

$$\tilde{u}(t, x, p) = \arg \max_{u \in U} \Big\{ p(t) \cdot f(t, x(t), u(t)) - L(t, x(t), u(t)) \Big\}.$$

Second we solve the **two-point boundary value problem** (TPBVP)

$$\begin{cases} \dot{x} = f(t, x, \tilde{u}(t, x, p)), \quad x(t_0) = x_0, \\ \\ \dot{p} = -p\frac{\partial f}{\partial x}(t, x, \tilde{u}(t, x, p)) + \frac{\partial L}{\partial x}(t, x, \tilde{u}(t, x, p)),\ p(T) = \nabla\psi(x(T)). \end{cases}$$

TPBVPs are commonly solved via so called *shooting methods* which consist in the following two steps iterations: i) guess an initial $\overline{p}$ and solve the Cauchy problem involving the above set of differential equations where the boundary condition at final time $p(T) = \nabla\psi(x(T))$ is replaced by the boundary condition at the initial time $p(t_0) = \overline{p}$; ii) Readjust $\overline{p}$ so to minimize $\Lambda(\overline{p}) := p(T) - \nabla\psi(x(T))$.

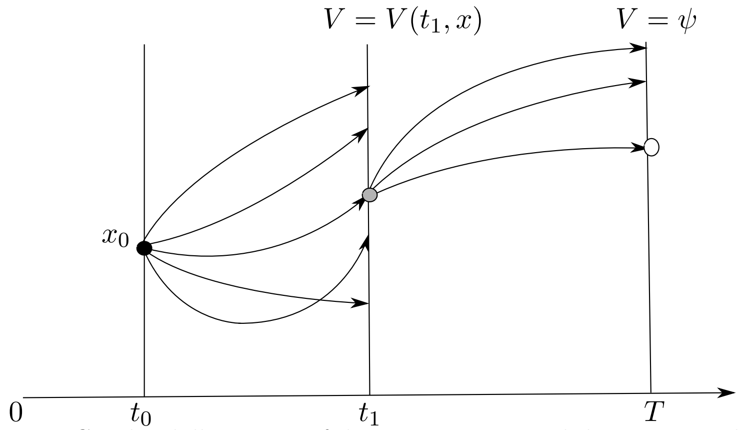

Figure 9.2: Graphical illustration of the HJB equation and the DP principle. Courtesy by Alberto Bressan, Noncooperative Differential Games. A Tutorial (2010).

9.1.3 Hamilton-Jacobi-Bellman (HJB)

A second seminal result made available by the American literature of the 50s is represented by the HJB equation. The HJB approach which is based on the dynamic programming (DP) principle is illustrated in Fig. 9.2. According to the DP principle, known as *Principle of Optimality*, "an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." (see [Bellman, 1957, Chap. III.3]). In other words, given an optimal trajectory any subtrajectory must be optimal as well. This allows us to decompose the original problem into infinite subproblems, one for each time t.

To do this, define the *value function* $V(t_0, x_0) := \sup_{u(\cdot)} J(u; t_0, x_0)$. From dynamic programming we have that the "value function today is the optimal running cost til tomorrow plus the value function tomorrow", which in formulas translates into:

$$\begin{aligned} V(t_0, x_0) = \sup\nolimits_{u(\cdot)} \Big\{ & V(t_1, x(t_1; t_0, x_0, u)) \\ & - \textstyle\int_{t_0}^{t_1} L(t, x(t; t_0, x_0, u), u(t)) dt \Big\}. \end{aligned} \tag{9.1}$$

By expanding according to Taylor $V(.)$ around the point (t_0, x_0), we obtain for $V(t_1, x(t_1))$ an expression in terms of $V(t_0, x_0)$ and the first-order derivatives of $V(.)$:

$$V(t_1, x(t_1)) = V(t_0, x_0) + \int_{t_0}^{t_1} \partial_t V(t, x(t)) + \nabla V \cdot f(t, x(t), u(t)).$$

Replacing $V(t_1, x(t_1))$ in (9.1) by its approximate expression above, we obtain the HJB equation in $[0, T] \times \mathbb{R}^m$:

$$\partial_t V(t, x) + \sup_{u \in U} \Big\{ \nabla V \cdot f(t, x, u) - L(t, x, u) \Big\} = 0.$$

9.2 Differential game

Consider a controlled dynamics, where $x \in \mathbb{R}^m$ is the state, u_i is the control of player $i = 1, 2$, and $U_i \subseteq \mathbb{R}^m$ is compact:

$$\dot{x}(t) = f(t, x(t), u_1(t), u_2(t)), \quad u_i(t) \in U_i.$$

Consider the optimization problem of player i:

$$\max_{u_i} J_i(u_1, u_2) := \psi_i(x(T)) - \int_0^T L_i(t, x(t), u_1(t), u_2(t)) dt.$$

We need to distinguish next between *open-loop* strategies where players know only the initial state x_0 and *Markovian strategies* (or closed-loop strategies) where players know the current state $x(t)$.

9.2.1 Open-loop Nash equilibrium

In the open-loop case, the strategies are functions of time (and initial state) only.

Definition 9.1. The pair $(u_1^*(t), u_2^*(t))$ is an NE if $u_i^*(\cdot)$ is a maximizer of the following cost functional problem:

$$\begin{cases} J_i(u_i, u_{-i}^*) := \psi_i(x(T)) - \int_0^T L_i(t, x(t), u_i(t), u_{-i}^*(t)) dt, \\ \\ x(0) = x_0, \quad \dot{x}(t) = f(t, x(t), u_i(t), u_{-i}^*(t)), \quad t \in [0, T]. \end{cases} \tag{9.2}$$

From PMP consider the following one-shot game: for every $(t, x) \in [0, T] \times \mathbb{R}^m$, and vectors $q_1, q_2 \in \mathbb{R}^m$,

$$\begin{aligned} \tilde{u}_1 &= \arg\max\nolimits_{\omega \in U_1} \{ q_1 \cdot f(t, x, \omega, \tilde{u}_2) - L_1(t, x, \omega, \tilde{u}_2) \}, \\ \\ \tilde{u}_2 &= \arg\max\nolimits_{\omega \in U_2} \{ q_2 \cdot f(t, x, \tilde{u}_1, \omega) - L_2(t, x, \tilde{u}_1, \omega) \}. \end{aligned} \tag{9.3}$$

Assume unique solution of the above problem (cf. Assumption **A2** on p. 26, [Bressan, 2010]). We need such an assumption for the corresponding map

$$(t, x, q_1, q_2) \mapsto \Big(\tilde{u}_1(t, x, q_1, q_2), \tilde{u}_2(t, x, q_1, q_2)\Big)$$

to be continuous.

If the pair $(\tilde{u}_1(t), \tilde{u}_2(t))$ is an NE then it solves the two-point boundary value problem

$$\begin{cases} \dot{x} = f(t, x, \tilde{u}_1, \tilde{u}_2), \quad x(t_0) = x_0, \\ \\ \dot{q}_1 = -q_1 \frac{\partial f}{\partial x}(t, x, \tilde{u}_1, \tilde{u}_2) + \frac{\partial L_1}{\partial x}(t, x, \tilde{u}_1, \tilde{u}_2), \; q_1(T) = \nabla \psi_1(x(T)), \\ \\ \dot{q}_2 = -q_2 \frac{\partial f}{\partial x}(t, x, \tilde{u}_1, \tilde{u}_2) + \frac{\partial L_2}{\partial x}(t, x, \tilde{u}_1, \tilde{u}_2), \; q_2(T) = \nabla \psi_2(x(T)). \end{cases} \tag{9.4}$$

The above condition is also sufficient if $x \mapsto H(t, x, q, \tilde{u}_{-i})$ and $x \mapsto \psi_i(x)$ are concave. The following example applies differential game techniques to a marketing competition scenario.

Example 9.1. Duopolistic competition

Two firms sell a same product in a same market. Variable $x_1(t) = x(t) \in [0, 1]$ is the market share of the first firm at time t while variable $x_2(t) = 1 - x(t)$ is the market share of the second firm. The control $u_i(t)$ represents the advertising effort of firm i at time t. Adopting the renowned *Lanchester* model we have that the market share of firm 1 evolves according to the following differential equation:

$$\dot{x}(t) = (1 - x)u_1 - xu_2, \quad x(0) = x_0 \in [0, 1].$$

Given the above dynamics, the ith firm chooses $t \mapsto u_i(t)$ in order to maximize

$$J_i = \int_0^T \left[a_i x_i(t) - c_i \frac{u_i^2(t)}{2} \right] dt + S_i x_i(T),$$

for suitable $a_i, c_i, S_i > 0$. In Step 1 the optimal controls are computed as functions of the adjoint variables:

$$\begin{cases} \tilde{u}_1(x, q_1, q_2) = \arg\max_{\omega \geq 0} \left\{ q_1 \cdot (1 - x)\omega - c_1 \frac{\omega^2}{2} \right\} = (1 - x)\frac{q_1}{c_1}, \\ \\ \tilde{u}_2(x, q_1, q_2) = \arg\max_{\omega \geq 0} \left\{ q_2 \cdot x\omega - c_2 \frac{\omega^2}{2} \right\} = x\frac{q_2}{c_2}. \end{cases}$$

In Step 2 one solves the two-point boundary value problem thus to obtain

$$\begin{cases} \dot{x} = (1-x)\tilde{u}_1 + x\tilde{u}_2 = (1-x)^2\frac{q_1}{c_1} + x^2\frac{q_2}{c_2}, \quad x(0) = x_0, \\ \dot{q}_1 = -q_1(\tilde{u}_1 + \tilde{u}_2) - a_1 = -q_1\left[(1-x)\frac{q_1}{c_1} + x\frac{q_2}{c_2}\right] - a_1,\ q_1(T) = S_1, \\ \dot{q}_2 = -q_2(\tilde{u}_1 + \tilde{u}_2) - a_2 = -q_2\left[(1-x)\frac{q_1}{c_1} + x\frac{q_2}{c_2}\right] - a_2,\ q_2(T) = S_2. \end{cases}$$

9.2.2 Closed-loop Nash equilibrium

In the closed-loop case, the strategies are functions of time and state.

Definition 9.2. The pair $(u_1^*(t,x), u_2^*(t,x))$ is an NE if $(t,x) \mapsto u_i^*(t,x)$ maximizes the following cost functional problem:

$$\begin{cases} J_i(u_i, u_{-i}^*(t,x)) := \psi_i(x(T)) - \int_0^T L_i(t, x(t), u_i(t), u_{-i}^*(t,x))dt, \\ x(0) = x_0, \quad \dot{x}(t) = f(t, x(t), u_i(t), u_{-i}^*(t,x)), \quad t \in [0,T]. \end{cases} \tag{9.5}$$

The computation of NE solutions is based on the solution of the corresponding HJB equations. Actually, we need to solve a system of PDEs in $[0,T] \times \mathbb{R}^m$ of the form:

$$\begin{cases} \partial_t V_1 + \nabla V_1 \cdot f(t, x, \tilde{u}_1, \tilde{u}_2) = L_1(t, x, \tilde{u}_1, \tilde{u}_2), \\ \partial_t V_2 + \nabla V_2 \cdot f(t, x, \tilde{u}_1, \tilde{u}_2) = L_2(t, x, \tilde{u}_1, \tilde{u}_2), \end{cases} \tag{9.6}$$

where, from PMP, $(\tilde{u}_1, \tilde{u}_2)$ solve the one-shot game: for every $(t,x) \in [0,T] \times \mathbb{R}^N$, and value functions $V_1, V_2 \in \mathbb{R}^N$,

$$\begin{aligned} \tilde{u}_1 &= \arg\max\nolimits_{\omega \in U_1}\{\nabla V_1 \cdot f(t, x, \omega, \tilde{u}_2) - L_1(t, x, \omega, \tilde{u}_2)\}, \\ \tilde{u}_2 &= \arg\max\nolimits_{\omega \in U_2}\{\nabla V_2 \cdot f(t, x, \tilde{u}_1, \omega) - L_2(t, x, \tilde{u}_1, \omega)\}. \end{aligned} \tag{9.7}$$

If the game is a zero-sum game (Rufus Isaacs, 1965) [Isaacs, 1965] then the above system of PDEs turns into a single one which is known as *Hamilton-Jacobi-Isaacs* equation:

$$\partial_t V_1 + \max_{\omega_1} \min_{\omega_2} \left\{ \nabla V_1 \cdot f(t, x, \omega_1, \omega_2) - L_1(t, x, \omega_1, \omega_2) \right\} = 0.$$

9.3 Linear-quadratic differential games

This class of games allows for the explicit computation of the optimal control policies. Consider a linear dynamics, where $x \in \mathbb{R}^m$ is the state, u_i is the control of player $i = 1, 2$, $U_i \equiv \mathbb{R}^{m_i}$ is compact:

$$\dot{x}(t) = A(t)x(t) + B_1(t)u_1(t) + B_2(t)u_2(t), \quad u_i(t) \in \mathbb{R}^{m_i}.$$

The problem boils down to the following optimization problem for player i:

$$\max_{u_i} J_i(u_1, u_2) := \psi_i(x(T)) - \int_0^T L_i(t, x(t), u_1(t), u_2(t))dt.$$

Assume that the terminal penalty is quadratic and of the form:

$$\psi_i(x(T)) = \frac{1}{2} x^T \overline{M}_i x,$$

and consider a quadratic running cost:

$$L_i(t, x(t), u_1(t), u_2(t)) = \frac{|u_i|^2}{2} + \frac{1}{2} x^T P_i x + \sum_{1,2} x^T Q_{ij} u_j.$$

Explicit computation of the optimal control policies $\tilde{u}_i$, can be performed as follows

$$\begin{array}{rcl} \tilde{u}_i(t, x, q_i) & = & \arg\max_{\omega \in \mathbb{R}^{m_i}} \left\{ q_i B_i(t)\omega - \frac{|\omega|^2}{2} - x^T Q_{ii}(t)\omega \right\} \\ & = & (q_i B_i(t) - x^T Q_{ii}(t))^T. \end{array} \tag{9.8}$$

Now, take $V_i(t, x) = \frac{1}{2} x^T M_i(t) x$ so that

$$\nabla V_i(t, x) = x^T M_i(t), \quad \partial_t V_i(t, x) = \frac{1}{2} x^T \dot{M}_i(t) x.$$

The HJB equation, which we rewrite here in compact form as $\partial_t V_i(t, x) = L_i - \nabla V_i \cdot f$, becomes

$$\begin{array}{l} \frac{1}{2} x^T \dot{M}_i(t) x \\ = \frac{1}{2}(x^T M_i B_i - x^T Q_{ii})(x^T M_i B_i - x^T Q_{ii})^T + \frac{1}{2} x^T P_i x \\ \qquad\qquad + \sum_{j=1,2} x^T Q_{ij}(x^T M_j B_j - x^T Q_{jj})^T \\ -x^T M_i (Ax + \sum_{j=1,2} B_j (x^T M_j B_j - x^T Q_{jj})^T). \end{array} \tag{9.9}$$

From the HJB we can derive the well known *Riccati differential equation* as follows. Observe that as the HJB below must hold for every x,

$$\begin{array}{l}\frac{1}{2}x^T\dot{M}_i(t)x \\ = \frac{1}{2}(x^T M_i B_i - x^T Q_{ii})(x^T M_i B_i - x^T Q_{ii})^T + \frac{1}{2}x^T P_i x \\ \qquad\qquad + \sum_{j=1,2} x^T Q_{ij}(x^T M_j B_j - x^T Q_{jj})^T \\ -x^T M_i(Ax + \sum_{j=1,2} B_j(x^T M_j B_j - x^T Q_{jj})^T).\end{array} \tag{9.10}$$

We can drop dependence on x which yields the following Riccati equation

$$\begin{array}{l}\frac{1}{2}\dot{M}_i(t) \\ = \frac{1}{2}(M_i B_i - Q_{ii})(M_i B_i - Q_{ii})^T + \frac{1}{2}P_i \\ \quad +\frac{1}{2}\sum_{j=1,2}[Q_{ij}(M_j B_j - Q_{jj})^T + (M_j B_j - Q_{jj})Q_{ij}^T] \\ -\frac{1}{2}(M_i A + A^T M_i) - \frac{1}{2}\sum_{j=1,2}[M_i B_j(M_j B_j - Q_{jj})^T \\ \qquad\qquad +(M_j B_j - Q_{jj})B_j^T M_i].\end{array} \tag{9.11}$$

9.4 Application: H^∞-optimal control

The problem is the one of a worst-case controller design already introduced in Section 2.3. Figure 9.3 illustrates the different blocks of the system.

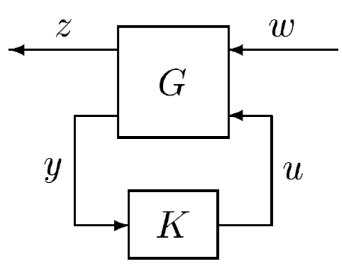

Figure 9.3: Block diagram of plant and feedback controller.

There the control is u, the disturbance is w, the controlled and measured outputs are z and y (all measurable in Hilbert spaces $\mathcal{H}_u, \mathcal{H}_w, \mathcal{H}_z, \mathcal{H}_y$) all satisfying the equation

$$\begin{cases} z = G_{11}(w) + G_{12}(u), \\ y = G_{21}(w) + G_{22}(u), \\ u = K(y). \end{cases} \tag{9.12}$$

A common assumption is that G_{ij} and controller $K \in \mathcal{K}$ are bounded causal linear operators and $\mathcal{K}$ is the controller space.

A main objective is the one of *disturbance attenuation*. Actually, for every fixed $K \in \mathcal{K}$, introduce bounded causal linear operators $T_K : \mathcal{H}_w \to \mathcal{H}_z$

$$T_K(w) = G_{11}(w) + G_{12}(I - KG_{22})^{-1}(KG_{21})(w).$$

We recall that causal means that the system is nonanticipative, namely, the output depends on past and current inputs but not on future inputs.

We wish to find the worst-case infimum of operator norm

$$\begin{cases} \inf_{K\in\mathcal{K}} \langle\langle T_K \rangle\rangle =: \gamma^*, \\ \langle\langle T_K \rangle\rangle = \sup_{w\in\mathcal{H}_w} \frac{\|T_K(w)\|_z}{\|w\|_w}. \end{cases} \tag{9.13}$$

The above problem turns into a TPZSG between controller and disturbance

$$\overbrace{\inf_{K\in\mathcal{K}} \sup_{w\in\mathcal{H}_w} \frac{\|T_K(w)\|_z}{\|w\|_w}}^{\text{upper bound}} \geq \overbrace{\sup_{w\in\mathcal{H}_w} \inf_{K\in\mathcal{K}} \frac{\|T_K(w)\|_z}{\|w\|_w}}^{\text{lower bound}}.$$

We can derive a so called *soft-contrained game*. To do this let γ^* be the attenuation level which satisfies

$$\inf_{K\in\mathcal{K}} \sup_{w\in\mathcal{H}_w} \|T_K(w)\|_z^2 - {\gamma^*}^2 \|w\|_w^2 \leq 0.$$

Define the parametrized cost (in $\gamma \geq 0$)

$$J_\gamma(K, w) := \|T_K(w)\|_z^2 - \gamma^2 \|w\|_w^2.$$

Then the problem is the one of finding the smallest value of $\gamma \geq 0$ under which the upper value is bounded (by zero).

We can then turn the problem into a linear quadratic zero-sum differential game. To do this, let the following state space representation be given

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) + D(t)w(t) \quad x(0) = x_0, \\ z(t) = H(t)x(t) + G(t)u(t), \\ y(t) = C(t)x(t) + E(t)w(t), \end{cases} \tag{9.14}$$

and for $\gamma \geq 0$ and $Q_T > 0$, consider the cost

$$L_\gamma(u,w) := x(T)^T Q_T x(T) + \int_0^T z(t)^T z(t)dt - \gamma^2 \int_0^T w(t)^T w(t)dt.$$

The zero-sum linear quadratic differential game appears as

$$\min_{u(\cdot)} \max_{w(\cdot)} L_\gamma(u,w).$$

It remains to solve it via Hamilton-Jacobi-Isaacs equations.

10

Stochastic games

This chapter builds upon the survey written by Eilon Solan on stochastic games in 2009 [Solan, 2009]. Similarly to differential games, stochastic games are games involving a state variable, but this evolves according to a controlled Markov chain rather than to a differential equation. In this sense, they represent an extension of Markov decision problems to the case with multiple interacting decision makers. This chapter skims through the foundations of stochastic games. In particular, after introducing the model in its generic form, we analyze seminal results for two-player stochastic zero-sum games and then we conclude by pointing out open questions. Further references on the topic are [Filar and Vrieze, 1996, Mertens and Neyman, 1981, Neyman and Sorin, 2003, Vieille, 2000].

10.1 Model

Stochastic games model the repeated interactions among players over a time horizon window in an environment that changes in response to the players' behaviors. The players' stage payoff depends on the players' current behaviors and on the environment, the latter described through a state variable. For a formal definition of a stochastic game let a set of players N be given, let t denote the time, and let S be the state space. The latter can be countable or uncountable, in which case it is supplemented with a σ-algebra of measurable sets. Let us introduce the set of actions $\mathcal{A}_i$ of player i and consider the valued function $A_i : S \to \mathcal{A}_i$ representing the available actions to player i in a given state. We can then introduce the set of action profiles as $SA := \{(s, a) : s \in S, a = (a_i)_{i\in N}, a_i \in A_i(s)\}$. In addition, for every player i, the stage payoff function is $u_i : SA \to \mathbb{R}$. We also denote the transition function by $q : SA \to \Delta(S)$ where $\Delta(S)$ is the space of probability distributions over S.

With the above model in mind we note that stochastic games constitute a generalization of i) games with *finite interactions*, if the play moves at time t to an absorbing state with payoff 0; ii) (static) *matrix game* if $t = 1$; iii) *repeated games* if we have one single state; iv) *stopping games* if the stage payoff is 0 until a player chooses *quit*, and the play moves to an absorbing state with nonnull payoff; v) *Markov decision problems* if we have just one single player.

In addition, actions determine current payoffs and future states and consequently also future payoffs. Furthermore, actions, payoffs, and transitions depend only on the current state.

10.1.1 Strategies

As in extensive games or differential games, the presence of a state variable induces the definition of a mapping from states to actions which we call *strategy*. In particular, given the past play at stage t, which we denote by

$$(s^1, a^1, s^2, a^2, \ldots, s^t),$$

where (s^k, a^k) is the action profile at time k, we call (pure) *stationary* strategy a strategy that depends only on the current state, that is:

$$\sigma_i(s^1, a^1, s^2, a^2, \ldots, s^t) \in A_i(s^t).$$

In other words, the past play $\sigma_i(s^1, a^1, s^2, a^2, \ldots, a^{t-1})$ does not count.

We can extend the definition to *mixed* strategies, which we indicate by

$$\sigma_i(s^1, a^1, s^2, a^2, \ldots, s^t) \in \Delta(A_i(s^t)),$$

where $\Delta(A_i(s^t))$ is the probability distribution on set $A_i(s^t)$.

In addition to this, we denote the space of stationary mixed strategies for player i by

$$X_i = \times_{s\in S}\Delta(A_i(s)).$$

Taking all players into account, the profile of mixed strategies is then

$$\sigma = (\sigma_i)_{i\in N}, \quad \sigma_i \in X_i.$$

Given that the game is played repeatedly over time, we need to introduce the space of infinite plays $H_\infty = SA^{\mathbb{N}}$ which is the set of all possible infinite sequences and is given by:

$$(s^1, a^1, s^2, a^2, \ldots, s^t, a^t, \ldots).$$

Now, let us observe that every profile of mixed strategies $\sigma = (\sigma_i)_{i\in N}$ and initial state s_1 induces a probability distribution $\mathbf{P}_{s_1,\sigma}$ on $H_\infty = SA^{\mathbb{N}}$. Given this, we are interested in the finite or infinite ($T \to \infty$) stream of payoffs

$$u_i(s^t, a^t), \quad t = 1, 2 \ldots, T.$$

10.1.2 Finite and infinite horizon

As in optimal control, we may have three different set-ups depending on whether we consider a finite or infinite horizon, and whether the players are myopic (shortsighted) or patient (farsighted). In particular, in the *finite horizon evaluation:* the interaction lasts exactly T stages. On the other hand, in the infinite horizon *discounted evaluation:* the interaction lasts many stages, and the players discount their stage payoffs in such a way that it is better to receive “1 dollar” today than the same dollar tomorrow. This is useful when we wish to capture a greedy or shortsighted attitude on the part of the players.

Still for the infinite horizon case, we can have the *limsup evaluation.* Here the interaction lasts many stages, and the players do not discount their stage payoffs, namely, the stage payoff at time t is insignificant if compared to the payoffs in all other stages. This fits the framework where the players are patient or farsighted and instantaneous payoffs' fluctuations do not count. For each of the above formulations we can define a payoff as detailed next. For the *T-stage payoff* we have

$$\gamma_i^T(s_1,\sigma) := \mathbb{E}_{s_1,\sigma}\Big[\frac{1}{T}\sum_{t=1}^{T} u_i(s^t,a^t)\Big].$$

Similarly for the *λ-discounted payoff* one obtains

$$\gamma_i^\lambda(s_1,\sigma) := \mathbb{E}_{s_1,\sigma}\Big[\lambda\sum_{t=1}^{\infty}(1-\lambda)^{t-1} u_i(s^t,a^t)\Big].$$

Finally, for the *limsup payoff* we have

$$\gamma_i^\infty(s_1,\sigma) := \mathbb{E}_{s_1,\sigma}\Big[\lim sup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} u_i(s^t,a^t)\Big].$$

In each of the three different scenarios we may wish to analyze *equilibria*. These have a different definition depending on which formulation we are thinking of as illustrated in the following definition.

Definition 10.1. We say that σ is a *T-stage ε-equilibrium* if

$$\gamma_i^T(s_1,\sigma) \geq \gamma_i^T(s_1,\sigma_i',\sigma_{-i}) - \varepsilon, \quad \forall\, s_1 \in S,\, i \in N,\, \sigma_i' \in X_i.$$

Differently, σ is a *λ-discounted ε-equilibrium* if

$$\gamma_i^\lambda(s_1,\sigma) \geq \gamma_i^\lambda(s_1,\sigma_i',\sigma_{-i}) - \varepsilon, \quad \forall\, s_1 \in S,\, i \in N,\, \sigma_i' \in X_i.$$

Finally, σ is a *limsup ε-equilibrium* if

$$\gamma_i^\infty(s_1,\sigma) \geq \gamma_i^\infty(s_1,\sigma_i',\sigma_{-i}) - \varepsilon, \quad \forall\, s_1 \in S,\, i \in N,\, \sigma_i' \in X_i.$$

It is worth noting that according to the above definitions, at the equilibrium the players benefit no more than ε from unilateral deviations.

10.2 Applications

Stochastic games arise in several application domains. We provide next a list, far from complete, of examples of applications that can be modeled through stochastic games.

Example 10.1. (Capital accumulation (Fishery)[Amir, 1996, Dutta and Sundaram, 1993, Levhari and Mirman, 1980, Nowak, 2003]) Two players jointly own a natural resource or productive asset. At every period they have to decide the amount of resource to consume. The amount that is not consumed grows by a known (or an unknown) fraction. The state is the current amount of resource. The action set is the amount of resource to be exploited in the current period. The transition between consecutive states is influenced by the decisions of all the players, as well as by the random growth of the resource.

Example 10.2. (Taxation [Chari and Kehoe, 1990, Phelan and Stacchetti, 2001]) A government sets a tax rate at every period. Each citizen decides at every period how much to work, and how much money to consume; the rest of the money grows by a known interest rate at the next period. The state is the citizens' amount of savings. The stage payoff of a citizen depends on the amount of i) money that he consumed, ii) free time he has, iii) the tax that the government collected in total.

The stage payoff of the government combines i) average stage payoff of citizens ii) amount of tax collected.

Example 10.3. (Communication network [Sagduyu and Ephremides, 2003]) A single-cell system with one receiver and multiple uplink transmitters share a single, slotted, synchronous classical collision channel. Transmitters at each time slot decide if and which packet to transmit. The state is the channel congestion. The stage payoff combines probability of successful transmission plus cost of transmission. The dropped packets are backlogged.

Example 10.4. (Queues [Altman, 2005]) Individuals may choose between private slow service provider, or powerful public service provider. The state is the current load of public and private service providers. The payoff is the time to be served.

10.3 Two player stochastic zero-sum game

Zero-sum games have since ever represented an important testbed for seminal results. This section focuses on zero-sum stochastic games and illustrates solution techniques based on dynamic programming on two classical examples.

In a two-player stochastic zero-sum game the sum of payoffs is zero, i.e.,

$$u_1(s,a) + u_2(s,a) = 0, \quad \forall (s,a) \in SA.$$

The game admits at most one equilibrium payoff (termed *value* of the game) at every initial state s_1. Each player's strategy σ_1 at an ε-equilibrium is ε-optimal in that it guarantees the value up to ε, i.e.,

$$\gamma_1^T(s_1, \sigma_1, \sigma_2) \geq \underbrace{v^T(s_1)}_{\text{value at } s_1} - \varepsilon, \quad \forall \sigma_2 \in X_2.$$

As for the existence results of equilibria provided by von Neumann for zero-sum static games, even for zero-sum stochastic games we have fundamental existence results given by Shapley in 1953 [Shapley, 1953], which were extended to multiplayer nonzero-sum games by Fink in 1964 [Fink, 1964].

Theorem 10.1. **[Shapley 1953]**: If all sets are finite, then for every λ there exists an equilibrium in stationary strategies.

Proof. Let us denote by $\mathcal{V}$ the space of all functions $v : S \to \mathbb{R}$. Define the zero-sum matrix game $G_s^\lambda(v)$ for all v. Let $A_1(s)$ and $A_2(s)$ be the space of actions at state s. The payoff which could be thought of as money that player 2 pays to player 1) is

$$\lambda u_1(s,a) + (1-\lambda) \sum_{s' \in S} q(s'|s,a) v(s').$$

Let us define the value operator $\phi_s(v) = val(G_s^\lambda(v))$. Non-expansiveness given by $\|\phi(v) - \phi(w)\|_\infty \leq (1-\lambda)\|v - w\|_\infty$ leads to a unique fixed point $\hat{v}^\lambda$. ☐

We say that the optimal mixed action σ_i in the matrix game $G_{s^t}^\lambda(\hat{v}^\lambda)$ is a λ-discounted 0-optimal strategy. The proof illustrated above is constructive and we will use it to solve the following two classical games.

10.4 Example 1: Big Match

The story behind the Big Match is very nice and simple [Vieille, 2000]. Once upon a time there was a king, who had to leave his reign for an undefined time. Before doing this he called in his trusted minister and put him in charge of his kingdom. The minister would not hear from the king until his return. If on that day, the king would find the minister at work, the king would recompense the minister by abducting in favor of him. On the other hand, if the minister would be caught idling, he would be imprisoned and tormented for ever and ever. The king had informers so he would be aware of how the minister behaved the past day.

A main question is now what the minister should do, given that the king pursues his own advantage. The minister knows that if he worked hard every day, then the king, being informed of this, would not come back, which would imply an everlasting miserable life. On the other hand, the minister knows that if he would decide not to work at all, the king would come very soon and the minister would be imprisoned.

Based on such a fairy tale, we can analyze the optimal strategy of both the king and the minister using a stochastic game. The game is described by the following matrices:

$$
\underset{\text{State } s_2}{\begin{array}{c|c|c|} & \text{L} & \text{R} \\ \cline{2-3} \text{T} & 0 \quad s_2 & 1 \quad s_2 \\ \cline{2-3} \text{B} & 1 \quad s_1 & 0 \quad s_0 \\ \cline{2-3} \end{array}}
\qquad
\underset{\text{State } s_1}{\begin{array}{c|c|} & \text{L} \\ \cline{2-2} \text{T} & 1 \quad s_1 \\ \cline{2-2} \end{array}}
\qquad
\underset{\text{State } s_0}{\begin{array}{c|c|} & \text{L} \\ \cline{2-2} \text{T} & 0 \quad s_0 \\ \cline{2-2} \end{array}}
$$

The king is the row player, and the minister the column player. For every day that the king does not come back (action T) the game transitions to the same state s_2 independently of what the minister does (say actions L or R). When the king returns (action B) then the game jumps to state s_1 (everlasting recompense) if the minister is hard working (action L) and to state s_0 (everlasting punishment) if the minister is found enjoying life (action R). Now, in order to solve the game and look for equilibria, let us apply the dynamic programming principle used in the proof of the theorem exposed earlier. For every

$v = (v_1, v_2, v_3) \in \mathcal{V} = \mathbb{R}^3$ the game $G^\lambda_{s_2}(v)$, $G^\lambda_{s_1}(v)$, and $G^\lambda_{s_0}(v)$ are given by

	L	R
T	$(1-\lambda)v_2$	$\lambda + (1-\lambda)v_2$
B	$\lambda + (1-\lambda)v_1$	$(1-\lambda)v_0$

Game $G^\lambda_{s_2}$

	L
T	$\lambda + (1-\lambda)v_1$

Game $G^\lambda_{s_1}$

	L
T	$(1-\lambda)v_0$

Game $G^\lambda_{s_0}$

Imposing the fixed point condition on states 0 and 1 we have

- $\hat{v}^\lambda_0 = val(G^\lambda_{s_0}(\hat{v}))$ which yieds $\hat{v}^\lambda_{s_0} = 0$
- $\hat{v}^\lambda_1 = val(G^\lambda_{s_1}(\hat{v}))$ which yieds $\hat{v}^\lambda_{s_1} = 1$.

Replacing the above values for state 2 we obtain

	L	R
T	$(1-\lambda)v_2$	$\lambda + (1-\lambda)v_2$
B	1	0

State s_2

From the indifference principle, the saddle-point of this game is obtained by solving

$$v_2 = y(1-\lambda)v_2 + (1-y)[\lambda + (1-\lambda)v_2] = y,$$
$$v_2 = x(1-\lambda)v_2 + (1-x) = x[\lambda + (1-\lambda)v_2],$$

where y is the probability that player 2 plays L and x the probability that player 1 plays T.

Then we obtain $\hat{v}^\lambda_2 = val(G^\lambda_{s_2}(\hat{v}))$ which yieds $\hat{v}^\lambda_{s_2} = \frac{1}{2}$. For the best response strategies we finally get

$$\sigma_2 = [\tfrac{1}{2}(L), \tfrac{1}{2}(R)], \quad \sigma_1 = [\tfrac{1}{1+\lambda}(T), \tfrac{\lambda}{1+\lambda}(B)].$$

In other words, the minister will have to work every two day on average, which means that every day he will toss a coin and depending on the result he will work hard or not. For the king the optimal strategy will depend on the discount factor, that is on how farsighted he is. In particular the probability of returning (action B) increases with the discount factor, that is, the more myopic the king is, the sooner he will come back. On the other hand, if the king is farsighted, the discount factor tends to zero or is small, and the probability of returning decreases to zero. It is worth noting that the discount factor influences the strategy of the king only and not of the minister. The main justification for this is that only the king can make the game jump to an absorbing state.

10.5 Example 2: Absorbing game

This example is a slight variation of the above one where also the minister's strategy can lead the game to an absorbing game and therefore we will see that the discount factor will affect the minister strategy as well. The game is described by the following matrices.

$$\begin{array}{c|c|c|} & L & R \\ \hline T & 0 \quad s_2 & 1 \quad s_1 \\ \hline B & 1 \quad s_1 & 0 \quad s_0 \\ \hline \end{array} \qquad \begin{array}{c|c|} & L \\ \hline T & 1 \quad s_1 \\ \hline \end{array} \qquad \begin{array}{c|c|} & L \\ \hline T & 0 \quad s_0 \\ \hline \end{array}$$

State s_2 State s_1 State s_0

The only difference with the Big Match is that now in (T, R) the game transitions to the absorbing state s_1. Applying the same technique as in the previous example, for every $v = (v_1, v_2, v_3) \in \mathcal{V} = \mathbb{R}^3$ the games $G^\lambda_{s_2}(v)$, $G^\lambda_{s_1}(v)$, and $G^\lambda_{s_0}(v)$ are given by

$$\begin{array}{c|c|c|} & L & R \\ \hline T & (1-\lambda)v_2 & \lambda + (1-\lambda)v_1 \\ \hline B & \lambda + (1-\lambda)v_1 & (1-\lambda)v_0 \\ \hline \end{array} \qquad \begin{array}{c|c|} & L \\ \hline T & \lambda + (1-\lambda)v_1 \\ \hline \end{array}$$

Game $G^\lambda_{s_2}$ Game $G^\lambda_{s_1}$

$$\begin{array}{c|c|} & L \\ \hline T & (1-\lambda)v_0 \\ \hline \end{array}$$

Game $G^\lambda_{s_0}$

Imposing the fixed point condition on the states we then have

- $\hat{v}^\lambda_0 = val(G^\lambda_{s_0}(\hat{v}))$ yieds $\hat{v}^\lambda_{s_0} = 0$,

- $\hat{v}_1^\lambda = val(G_{s_1}^\lambda(\hat{v}))$ yieds $\hat{v}_{s_1}^\lambda = 1$,
- $\hat{v}_2^\lambda = val(G_{s_2}^\lambda(\hat{v}))$ yieds $\hat{v}_{s_2}^\lambda = \frac{1-\sqrt{\lambda}}{1-\lambda}$.

The latter equation is obtained from

$$\begin{aligned} v_2 &= y(1-\lambda)v_2 + (1-y) = y, \\ v_2 &= x(1-\lambda)v_2 + (1-x) = x, \end{aligned}$$

where y is the probability that player 2 plays L and x the probability that player 1 plays T.

The resulting best response strategies are then

$$\sigma_2 = [\tfrac{1-\sqrt{\lambda}}{1-\lambda}(L), \tfrac{\sqrt{\lambda}-\lambda}{1-\lambda}(R)], \quad \sigma_1 = [\tfrac{1-\sqrt{\lambda}}{1-\lambda}(T), \tfrac{\sqrt{\lambda}-\lambda}{1-\lambda}(B)].$$

As expected both the minister's and the king's strategies depend on the discount factor.

10.6 Results and open questions

A seminal result for two-player zero-sum games is the one provided by Mertens and Neyman in 1981 [Mertens and Neyman, 1981] about the existence of a uniform equilibrium, which we report next.

Theorem 10.2. **[Mertens and Neyman 1981, [Mertens and Neyman, 1981]]** For two-player zero-sum games each player has a strategy that is ε-optimal for every discount factor sufficiently small.

The above result was extended to non zero-sum games by Vieille in 2000 thus leading to the following theorem.

Theorem 10.3. **[Vieille 2000, [Vieille, 2000]]** For every two-player non zero-sum stochastic game there is a strategy profile that is an ε-equilibrium for every discount factor sufficiently small.

Many open questions are still available. In particular, a main issue is to find for every stochastic game a strategy profile that is an ε-equilibrium for every discount factor sufficiently small. Other open problems involve the identification of classes of games where one has a simple strategy profile that is an ε-equilibrium for every discount factor sufficiently small, (e.g. stationary strategy, periodic strategy). Relevant aspects involve also the numerical analysis and the algorithms for the computation of equilibria whenever explicit solutions are not available. In this sense, for two-player zero-sum games most algorithms are based on linear programming. Such algorithms have also been extended to non zero-sum game (see for instance the well-known *Lemke-Howson algorithm*). Other algorithms are based on fictitious play, value iterates, and policy improvement.

Future directions may involve i) the approximation of games with infinite state and action spaces by finite games; ii) The formulation and study of stochastic games in continuous time; iii) The existence of a uniform equilibrium and a limsup equilibrium in multi-player stochastic games with finite state spaces and action spaces; iv) The development of efficient algorithms that calculate the value of two-player zero-sum games; v) the study of approachable and excludable sets in stochastic games with vector payoffs. This last topic, together with the notion of approachability and attainability is the subject of the next chapter.

11

Games with vector payoffs

This chapter illustrates concepts developed within the context of *games with vector payoffs*. The foundations of games with vector payoffs were laid by David Blackwell in the 50s in his seminal paper on *approachability* [Blackwell, 1956]. Blackwell's result is brilliantly recalled in the book by Cesa-Bianchi and Lugosi [Cesa-Bianchi and Lugosi, 2006]. Vector payoffs arise whenever the outcome of a game involves multiple noninterchangeable items. An extremely simple case is a job interview in which the potential employer and the candidate negotiate different aspects of the contract, from salary to career perspectives, from benefits to days off. If the interaction between the players occurs repeatedly in time thus generating a stream of instantaneous payoffs, players may be interested in controlling the evolution of such a stream and in particular of the time averaged payoff in order to make it "approach" a predetermined set. Conditions for this to happen are collected in the literature under the umbrella of approachability. Attainability extends the theory to the case where the controlled variable is the cumulative rather than the average payoff [Bauso et al., 2014, Lehrer et al., 2011].

11.1 An illustrative simple game

Two players play a discrete-time repeated game and corresponding to any action profile is a 2-dimensional vector payoff as displayed in the following bimatrix:

$$\begin{pmatrix} (6,7) & (1,7) & (6,2) & (1,2) \\ (6,-4) & (1,-4) & (6,-9) & (1,-9) \\ (-3,-1) & (-8,-1) & (-3,-6) & (-8,-6) \\ (-3,10) & (-8,10) & (-3,5) & (-8,5) \end{pmatrix}.$$

Consider the total payoff up to stage t and denote it by $x(t)$ and let us indicate the average payoff by $\bar{x}(t)$. Thus, at time $t = 1$, assuming that the action profile is (Top,Left) we have that the total payoff is equal to the average payoff, that is $x(1) = \bar{x}(1) = (6,7)$ (see the red block in the bimatrix below),

$$\begin{pmatrix} \boxed{(6,7)} & (1,7) & (6,2) & (1,2) \\ (6,-4) & (1,-4) & (6,-9) & (1,-9) \\ (-3,-1) & (-8,-1) & (-3,-6) & (-8,-6) \\ (-3,10) & (-8,10) & (-3,5) & (-8,5) \end{pmatrix}.$$

Now, at time $t = 2$, suppose that the action profile is (Bottom,Right), then we obtain that the total payoff is $x(2) = (-2,12)$ whereas the average payoff is $\bar{x}(2) = (-1,6)$ (see the red block in the bimatrix below),

$$\begin{pmatrix} (6,7) & (1,7) & (6,2) & (1,2) \\ (6,-4) & (1,-4) & (6,-9) & (1,-9) \\ (-3,-1) & (-8,-1) & (-3,-6) & (-8,-6) \\ (-3,10) & (-8,10) & (-3,5) & \boxed{(-8,5)} \end{pmatrix}.$$

If at time $t = 3$ the action profile were (Top,Left) again we would obtain that the total payoff is $x(3) = (4,19)$ whereas the average payoff is $\bar{x}(3) = (\frac{4}{3}, \frac{19}{3})$ and we could continue iterating the process at time

4, 5 . . . and so forth (see again the red block in the bimatrix below),

$$\begin{pmatrix} \boxed{(6,7)} & (1,7) & (6,2) & (1,2) \\ (6,-4) & (1,-4) & (6,-9) & (1,-9) \\ (-3,-1) & (-8,-1) & (-3,-6) & (-8,-6) \\ (-3,10) & (-8,10) & (-3,5) & (-8,5) \end{pmatrix}.$$

11.1.1 What do we know about vector payoffs?

Dating back to the 1950s, the literature provides convergence conditions in games with vector payoffs. The main result were given by David Blackwell in his seminal paper [Blackwell, 1956]:

Blackwell, D. (1956b) "An Analog of the MinMax Theorem For Vector Payoffs," Pacific J. of Math., 6, 1-8.

It was Blackwell who first formalized the concept of "approachable set" which we copy below.

Definition 11.1. A set of payoff vectors A is *approachable* by P_1 if she has a strategy such that the average payoff up to stage t, $\bar{x}(t) := \frac{x(t)}{t}$, converges to A, regardless of the strategy of P_2.

	L	R
T	$(0,0)$	$(0,0)$
B	$(1,1)$	$(1,0)$

Figure 11.1: Approachability example provided in the book by Mashler, Solan, and Zamir [Maschler et al., 2013].

Example 11.1. Consider the bimatrix given in Fig. 11.1, which is borrowed from the book by Mashler, Solan, and Zamir [Maschler et al., 2013].

It can be shown that sets C_1, C_2, and C_3 shown in Fig. 11.2 are all approachable. Set C_1 is the singleton $(0,0)$ and is approachable as player 1 can always stick to action T which yields the vector payoff

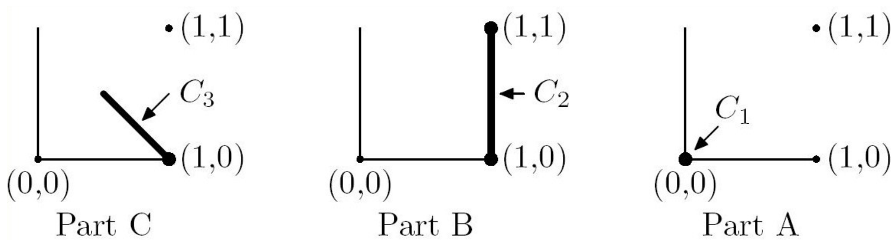

Figure 11.2: Example of approachable sets (C_1, C_2, and C_3) as reported in the book by Mashler, Solan, and Zamir [Maschler et al., 2013].

$(0,0)$. Set C_2 is also approachable as player 1 can select action B at every time which returns the vector payoff $(1, \$)$ where the second component is in the interval $[0,1]$ whatever player 2 does. Finally, set C_3 is also approachable given that player 1 can play the strategy

$$\begin{cases} B & \text{if } \bar{x}_1(t-1) + \bar{x}_2(t-1) < 1 \\ T & \text{otherwise.} \end{cases}$$

11.2 A geometric insight into approachability

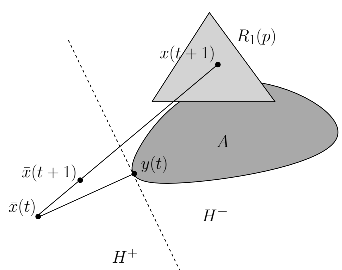

Figure 11.3: Geometric illustration of Blackwell's Approachability Principle.

Figure 11.3 provides a geometric illustration of Blackwell's Approachability Principle. Let set A be the set we wish to approach and denote by $y(t)$ the projection of the current average payoff $\bar{x}(t)$ onto A. Consider the supporting hyperplane (dashed line) for A in $y(t)$, which is given by

$$H = \{z \in \mathbb{R}^d | \, \langle z - y(t), \bar{x}(t) - y(t) \rangle = 0\}.$$

Furthermore let H^+ and H^- be the corresponding positive and negative half-spaces, namely

$$H^+ = \{z \in \mathbb{R}^d | \langle z - y(t), \bar{x}(t) - y(t) \rangle \geq 0\},$$
$$H^- = \{z \in \mathbb{R}^d | \langle z - y(t), \bar{x}(t) - y(t) \rangle \leq 0\}.$$

The approachability principle can be stated as follows:

Theorem 11.1. (Blackwell's Approachability Principle) Set A is approachable if for any $\bar{x}(t) \in H^-$, $\exists\ p$ such that $R_1(p) \subset H^+$, where $R_1(p)$ is the set of payoffs when player 1 sticks to the mixed action p and for all possible actions of player 2.

11.3 More recent results

The literature on approachability has developed in different directions. A first one involves the extension of the approachability conditions to infinite dimensional spaces, that is, when the payoff vector turns into an infinite dimensional signal. The analysis requires the adaptation of concepts like distance introduced for finite Eucleadian spaces to infinite dimensional spaces[Lehrer, 2002]

A second major development sheds light on connections with differential games [Soulaimani et al., 2009] (see also [Lehrer and Sorin, 2007]). In [Soulaimani et al., 2009] the authors show that a classical approachability problem can be turned into a zero-sum differential game. This is possible once we describe the average payoff evolution through a controlled differential equation subject to controlled inputs (the strategy of player 1) and adversarial disturbances (the strategy of player 2). After some manipulation the system takes on the form of an uncertain dynamical system with multiplicative uncertainty and the approachability question then changes into a standard reachability control problem.

A third line of research investigates on the role of approachability condition in *regret minimization* [Hart, 2005, Hart and Mas-Colell, 2001, 2003]. The latter is a topic concerned with learning from past mistakes based on a posteriori observations. Every player builds its strategy on a so-called *regret vector*, which represents the benefit if the player had constantly played another alternative strategy rather than the current one. At a Nash equilibrium the regret vector of all players is nonpositive. Thus, converging to a Nash equilibrium corresponds to driving all regret vectors to the negative orthant on the part of the players.

11.4 Attainability

In a repeated game, attainability refers to convergence of the cumulative payoffs rather than average payoff [Lehrer et al., 2011]. This is captured by the following definition.

Definition 11.2. (Attainability, [Bauso et al., 2014, Lehrer et al., 2011]) A set of payoff vectors A is *attainable* by P_1 if she has a strategy such that the *total payoff* up to stage t, $x(t)$, "converges" to A, regardless of the strategy of P_2.

This apparently subtle distinction with approachability has however profound implications as we will detail in the following.

11.5 Attainability and robust control

It was the network robust control problem reported below that paved the way to the definition of attainability [Bauso et al., 2006, 2010].

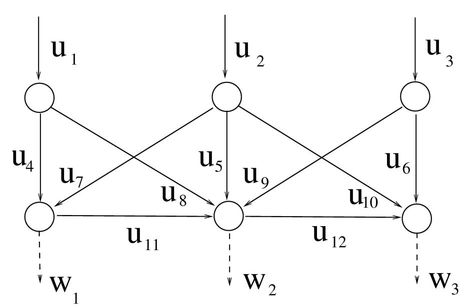

Figure 11.4: Network robust control problem.

Consider the network in Fig. 11.4 involving *uncontrolled* flows/demand $w(t) \in \mathcal{W}, \quad \forall t$, and *controlled* flows/supply $u(t) \in \mathcal{U}, \quad \forall t$. The buffer (excess supply) at the nodes evolves with time according to the dynamics:

$$\dot{x}(t) = Bu(t) - w(t),\, x(0) = \zeta,$$

where B is the incidence matrix of the network and ζ is the initial state (initial configuration of excesses). Essentially the above differential equation identifies the discrepancy between the incoming flow to and the outgoing flow from each node.

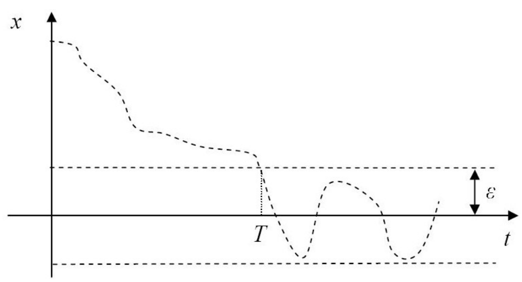

Figure 11.5: Tube reachability and robustness.

The problem consists in designing a state feedback control to drive the state of the system towards the origin. Robustness is intended as the ability to do this without knowing the realization of the uncontrolled flows w but only the set it belongs to. A disturbance for which only the bounding set is known is referred to as *unknown but bounded* [Bertsekas and Rhodes, 1971].

A radically different perspective on this problem is provided by the theory of repeated games with vector payoffs. In this perspective we assume that the controlled flows are selected by P_1 who plays $u(t)$ and the uncontrolled flow are chosen by P_2 who plays w. Thus, $\dot{x}(t)$ is the instantaneous vector payoff of the game and $x(t)$ is the total payoff up to time t.

Example 11.2. This example should clarify how to turn a robust control problem into an attainability problem.

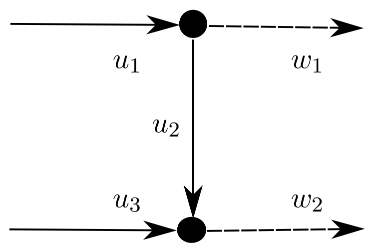

Figure 11.6: An example of a network flow control problem turned into an attainability problem.

Given the network in Fig. 11.6, suppose that the controlled flows can be processed in batches, namely,

$$u(t) \in \left\{ \begin{bmatrix} 1 \\ -2 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \begin{bmatrix} -5 \\ 1 \\ -5 \end{bmatrix}, \begin{bmatrix} -5 \\ 1 \\ 6 \end{bmatrix} \right\}.$$

Likewise, for the uncontrolled flows we have

$$w(t) \in \left\{ \begin{bmatrix} -3 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right\}.$$

In other words, both flows, controlled and uncontrolled, take values in predetermined discrete sets, which we now reinterpret as discrete action sets. Then, the dynamics for the excesses takes the form:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix} - \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}.$$

It is straightforward to note that once we substitute for $u(t)$ and $w(t)$ the values in the discrete sets we end up with the bimatrix below, which is exactly the one we presented at the beginning of the chapter:

$$\begin{pmatrix} (6,7) & (1,7) & (6,2) & (1,2) \\ (6,-4) & (1,-4) & (6,-9) & (1,-9) \\ (-3,-1) & (-8,-1) & (-3,-6) & (-8,-6) \\ (-3,10) & (-8,10) & (-3,5) & (-8,5) \end{pmatrix}.$$

Each entry represents the 2-dimension payoff corresponding to any possible action profile. For instance, the entry $(6,7)$ in the first row -

first column results from replacing $u(t) = \begin{bmatrix} 1 \\ -2 \\ 6 \end{bmatrix}$ and $w(t) = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$ and so forth.

11.6 Model and results

This section expands on the model and the results available so far on attainability. Let a two-player repeated game $(A_1, A_2, \mathbf{g})$ be given where A_i is the action space of P_i and $\mathbf{g} : A_1 \times A_2 \to [-1, 1]^d$ is the d-dimensional payoff. Denote by $(a_i^t)_{t \in \mathbb{R}_+}$ a *non-anticipative behavior strategy* for player i characterized by the following facts:

- $(a_i^t)_{t \in \mathbb{R}_+}$ takes values in $\Delta(A_i)$
- $\exists$ an increasing sequence of times $\tau_i^1 < \tau_i^2 < \tau_i^3 < \dots$ such that a_i^t is measurable with respect to the information available at τ_i^k, $\tau_i^k \leq t < \tau_i^{k+1}$.

A possible non-anticipative behavior strategy is illustrated in Fig. 11.7. At time τ_i^0 (the origin of the axes) player i sets the next time τ_i^1 and chooses to play the mixed strategy $(\frac{1}{2}, \frac{1}{2})$, i.e., he chooses T or B with equal probability and sticks to such a strategy for the whole interval from 0 to τ_i^1. At time τ_i^1 player i gets new information about the history of the system (past play of his opponent), sets a new successive time τ_i^2 and a new mixed strategy $(1, 0)$, that is, always T within the interval $[\tau_i^1, \tau_i^2)$. At time τ_i^2, based on the updated information now available, player i, sets a new successive time τ_i^3 and a new mixed strategy $(\frac{1}{3}, \frac{2}{3})$, that is, T with probability $\frac{1}{3}$ and B with probability $\frac{2}{3}$, and maintains this choice all over the interval $[\tau_i^2, \tau_i^3)$ and so forth.

The set-up is complete if we add to it the payoff at time t given the mixed actions of the players, which we denote g_t, and its integral (which represents the cumulative payoff) $x(t) = \int_{\tau=0}^{t} g_\tau(\text{mixed action pairs at time } \tau) d\tau$.

We are then ready to formalize the notion of attainable set as provided below.

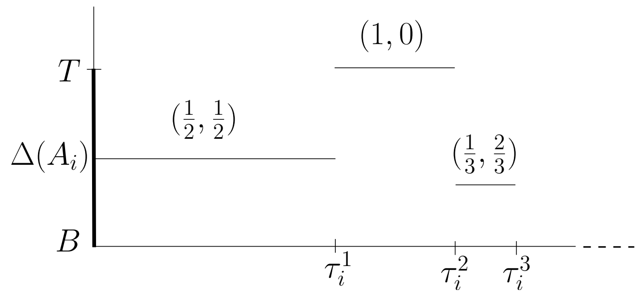

Figure 11.7: Non-anticipative strategy for player i: $(\frac{1}{2}, \frac{1}{2})$ in the first interval, $(1, 0)$ in the second interval, and $(\frac{1}{3}, \frac{2}{3})$ in the third interval.

Definition 11.3. A set A in $\mathbb{R}^d$ is *attainable* by P_1 if there is $T > 0$ such that for every $\epsilon > 0$ there is a strategy σ_1 of $P1$ such that

$$dist(x(t)[\sigma_1, \sigma_2], A) \leq \epsilon, \quad \forall t \geq T, \forall \sigma_2.$$

The above definition is further illustrated in geometric terms in Fig. 11.8. Let A be the set we wish to attain, $B(A, \epsilon)$ is the set of

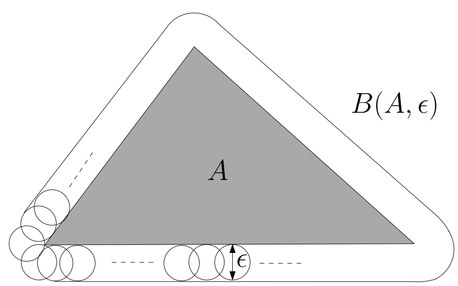

Figure 11.8: Epsilon ball of attainable set A.

points that are far from A no more than ϵ, that is

$$B(A, \epsilon) := \{z : dist(z, A) \leq \epsilon\}.$$

In the following we present three main results related to attainability of zero, attainability of a specific point different from zero, and attainability of every point in the space of the payoffs. In all of the following results a central role is played by the projected game, and its value.

More specifically, we will see that the value of the projected game, denoted by v_λ, must be bounded in sign, namely, $v_\lambda > 0$ or $v_\lambda \geq 0$. To understand this aspect, let a game with vector payoffs be given as represented by the bimatrix below (left):

$$\begin{pmatrix} (\#,\#) & (\#,\#) \\ (\#,\#) & (\#,\#) \end{pmatrix} \Rightarrow \begin{pmatrix} \langle \lambda, (\#,\#) \rangle & \langle \lambda, (\#,\#) \rangle \\ \langle \lambda, (\#,\#) \rangle & \langle \lambda, (\#,\#) \rangle \end{pmatrix}.$$

We can derive a matrix game by premultiplying all entries by a predetermined vector $\lambda \in \mathbb{R}^d$ (think of this as a direction in the space of the payoffs) and obtain the matrix game on the right. For the resulting game, which is now a two-player zero-sum game with scalar payoff, we can then compute the equilibrium payoff v_λ, which we recall to be the *value* of the game. The value depends on the direction of projection and therefore we use the index λ.

Given the above definition of value of the projected (along direction λ) game, we are ready to state the main results of attainability.

The first result addresses conditions for attainability of vector $\vec{0}$.

Theorem 11.2. The following conditions are equivalent.

B1 Vector $\vec{0} \in \mathbb{R}^d$ is attainable by P_1;

B2 $v_\lambda \geq 0$ for every $\lambda \in \mathbb{R}^d$.

The interpretation of the above theorem is that for the origin of the payoff space to be attainable one needs that the value of every projected game be bounded in sign.

Condition ***B1*** is in spirit very similar to the Blackwell's approachability condition. This is shown in Fig. 11.9. The case illustrated here shows the condition when the attainable set includes only point zero. Given the cumulative payoff $x(\tau_1^k)$, after projection onto zero we identify the direction $\lambda = -\frac{1}{\|x(\tau_1^k)\|}x(\tau_1^k)$. For any possible payoff in the set $R_1(p)$ a priori given, which is obtained when player 1 adopts strategy p, the inner product between the payoff and the direction λ is nonnegative, that is, the payoff is confined within the positive half space H^+ as well as λ.

Such a condition is necessary also to the attainability of a specific point in the space of payoffs different from zero as established in the next theorem.

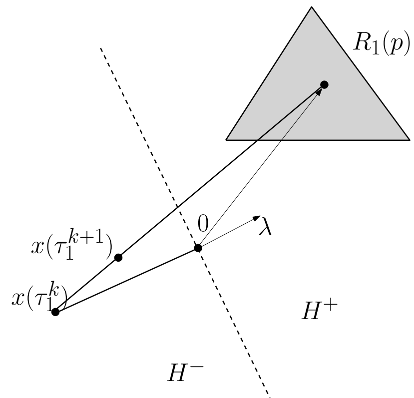

Figure 11.9: Geometric illustration of the attainability conditions.

Theorem 11.3. Vector $z \in \mathbb{R}^d$ ($\neq \vec{0}$) is attainable by $P_1 \Leftrightarrow$

B1 The vector $\vec{0} \in \mathbb{R}^d$ is attainable by P_1 and

B3 for every function $f : \Delta(A_1) \to \Delta(A_2)$, vector z is in

$$Cone(f) := \Big\{y \in \mathbb{R}^d |\, y = \sum_{p \in A_1} \alpha_p g(p, f(p)) \colon \alpha_p \geq 0\, \forall p\Big\}.$$

Rather than a formal proof, we provide next a geometric illustration of the underlying idea behind the proof. In particular, to see why ***B3*** is relevant for the attainability of z suppose that P_1 can play the mixed action $p \in \Delta(\{T, B\})$, where $\Delta(\{T, B\})$ is the set of probability distributions over $\{T, B\}$, and P_2 plays $f(p)$. Then the payoff $x(\tau_1^1)$ belongs to segment ab as in Fig. 11.10. Now, suppose that P_1 plays B in the interval $0 \leq t \leq \tau_1^1$, then the payoff $x(\tau_1^1)$ is the extreme point of the segment in red in Fig. 11.11 and, given all possible mixed strategies in the second interval the payoff $x(\tau_1^2)$ belongs to segment cd.

Differently from the previous interval, suppose that P_1 plays T in the interval $\tau_1^1 \leq t \leq \tau_1^2$, then the payoff $x(\tau_1^2)$ is the extreme point of the new segment in red in Fig. 11.12 and, given all possible mixed strategies in the third interval the payoff $x(\tau_1^3)$ belongs to segment ef. At time τ_1^3, assuming that P_1 plays a mixed strategy $(\frac{1}{2}, \frac{1}{2})$ in the interval $\tau_1^2 \leq t \leq \tau_1^3$ the resulting payoff $x(\tau_1^3)$ is the extreme point of the third segment in red in Fig. 11.13. Now, given that $x(\tau_1^3)$ is approxi-

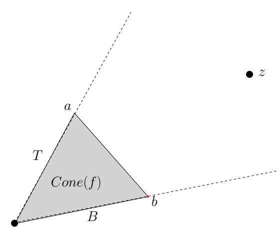

Figure 11.10: Geometric illustration of Theorem 11.3: time τ_1^1.

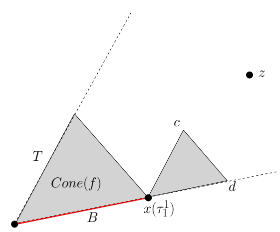

Figure 11.11: Geometric illustration of Theorem 11.3: time τ_1^2.

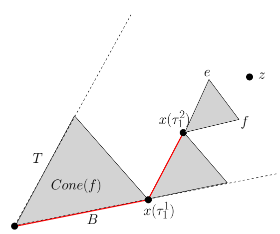

Figure 11.12: Geometric illustration of Theorem 11.3: time τ_1^3.

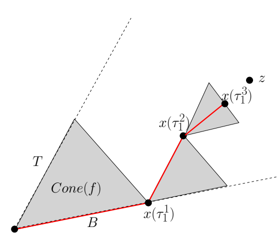

Figure 11.13: Geometric illustration of Theorem 11.3: from time τ_1^3 on.

mately close to point z, which we wish to attain, it is straightforward that in the game from τ_1^3 to ∞ the total payoff has to be approximately close to zero and this is true only if $\vec{0}$ is attainable. The very same way in which we have built the trajectory of the cumulative payoff explains why all possible trajectories $\{x(\tau_1^k)\}_{k=0,\dots,\infty}$ are within $Cone(f)$.

The last result in which we comment here shows that if the sign of the value of the projected game is strictly positive for every direction $\lambda \in \mathbb{R}^d$ then every vector in the payoff space is attainable.

Theorem 11.4. The following statements are equivalent:

C1 $v_\lambda > 0$ for every $\lambda \in \mathbb{R}^d$;

C2 Every vector $z \in \mathbb{R}^d$ is attainable by player 1.

11.7 Conclusions

This chapter has provided a perspective on repeated games with vector payoffs and introduced two main concepts, approachability first introduced by D. Blackwell in 1956 and attainability coined more recently in [Bauso et al., 2014, Lehrer et al., 2011]. It has been shown that while approachability looks at the time averaged payoff, attainability explores convergence conditions for the total or cumulative payoff. There are still many questions open to future analysis such as the characterization of attainable sets, the identification of conditions in the case of discounted payoff, and the study of approachability in evolutionary games

12

Mean–field games

The mean–field theory of dynamical games with large but finite populations of asymptotically negligible agents (as the population size goes to infinity) originated in the work of M.Y. Huang, P. E. Caines and R. Malhamé [Huang et al., 2003, 2006, 2007] and independently in that of J. M. Lasry and P.L. Lions [Lasry and Lions, 2006a,b, 2007], where the now standard terminology of Mean Field Games (MFG) was introduced. In addition to this, the closely related notion of *oblivious equilibria* for large population dynamic games was introduced by G. Weintraub, C. Benkard, and B. Van Roy [Weintraub et al., 2005] in the framework of Markov Decision Processes. The theory of mean–field games builds upon the notion of nonatomic player introduced first by by Robert Aumann for a continuum of traders [Aumann, 1964] and successively by Jovanovich and Rosenthal in a sequential game [Jovanovic and Rosenthal, 1988]. Mean field games arise in several applicative domains such as economics, physics, biology, and network engineering (see [Achdou et al., 2012, Bauso and Pesenti, 2013, Bagagiolo and Bauso, 2014, Gueant et al., 2010, Huang et al., 2007, Lachapelle et al., 2010, Zhu and Başar, 2011]).

This chapter provides an overview on the theory of games with many negligible agents. After presenting the main set-up we shall discuss a few stylized examples borrowed from [Gueant et al., 2010]. The last part skims through some available results on existence and uniqueness of solutions, linear quadratic mean–field games, and robustness.

12.1 The classical model

Consider a number $N \to \infty$ of homogeneous agents or players. "Homogeneous" means that every player in a same state $x \in \mathbb{R}^n$ will adopt the same state-dependent strategy $u(x(t), t)$. Assume that every player's state evolves according to the first-order differential equation:

$$\dot{x}(t) = u(x(t), t), \quad x_0 \in \mathbb{R}^n.$$

Actually, $u(x, t)$ defines a vector field in $\mathbb{R}^n$. Thus if the players are salt particles flowing through a river bed, we can think of $x(t) \in \mathbb{R}^2$ as the geographical position of the particle at time t, and $u(x(t), t)$ as its velocity. The state could also represent an opinion. The state $x \in [0, 1]^2$ would then be the opinion on how good is the indian restaurant and the chinese restaurant. The vertex top right of $x \in [0, 1]^2$ would then represent the opinion $(1, 1)$ which means that both restaurants are excellent, while the vertex bottom left is $(0, 0)$ saying that both restaurant are bad and so forth. The vector field would then represent how opinions evolve with time.

Back to the example of salt particles, assume that we wish to describe the concentration or density of particles in a given point in space x and time t. We can do this by using a density function $m(x, t)$ depending on both space x and time t. When such a scalar function is immersed in a vector field, its value evolves according to the so-called *advection equation*, which is given by

$$\partial_t m + div(m \cdot u(x, t)) = 0, \text{ in } \mathbb{R}^n \times [0, T].$$

Essentially, the advection equation, also known as *transport equation*, is a conservation law establishing that the partial derivative with respect to time of the density balances the divergence of the scalar m immersed in the vector field. To see why this is a conservation law, let us focus on point x in Fig. 12.1 and suppose that it is a “source”, namely there is an outgoing flow from it. As the divergence represents the flow traversing the surface around x in the limit when the radius tends to zero, then an outgoing flow would correspond to a divergence term positive in sign. Then the partial derivative shall be negative which means that the concentration of salt particles decreases and this is in fit with what we expect to happen. On the contrary, if x is a “sink”, then the divergence term is negative, the partial derivative is positive and the concentration increases.

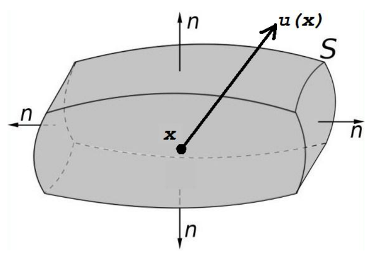

Figure 12.1: Physical interpretation of the divergence operator used in the advection equation.

Now, suppose that the particles are smart, in the sense that they wish to minimize a cost functional of the form

$$\int_0^T \Big[\underbrace{\frac{1}{2}|u(x(t),t)|^2}_{\text{penalty on control}} + \underbrace{g(x(t),m(\cdot,t))}_{\text{...on state \& distribution}} \Big] dt + \underbrace{G(x(T),m(\cdot,T))}_{\text{...on final state}}.$$

It is worth noticing that the above functional has the classical structure of an optimal control functional except for the presence of the term $m(\cdot)$ in the integrand and in the terminal penalty. From optimal control we have that the optimal feedback control is given by

$$u(x(t),t) = -\nabla_x J(x(t),t),$$

where $J(.,.)$ is the so called value function or optimal cost, which must be obtained by solving the famous Hamilton-Jacobi-Bellman (HJB) equation. We shall explain later on how we obtain such an equation.

Thus, a classical mean–field game model looks like two coupled partial differential equations (PDEs) in $\mathbb{R}^n \times [0,T]$:

$$\begin{array}{ccc} -\partial_t J + \frac{1}{2}|\nabla_x J|^2 = g(x,m) & & \text{(HJB) - backward} \\ u \downarrow \qquad\qquad \uparrow m & & \\ \partial_t m + div(m \cdot u(x)) = 0 & & \text{(advection) - forward} \end{array}$$

The above system of PDEs must be solved using the boundary conditions $m(\cdot,0) = m_0$, and $J(x,T) = G(x, m(\cdot,T))$. In other words, given the finite horizon formulation provided so far, the first PDE is to be solved backwards in time with boundary condition on the cost at time T. Such equation is parametrized in the density $m(\cdot)$ and is solved in the variable *value function* $J(\cdot)$ from which we obtain the optimal control $u(\cdot)$. It can be interpreted as returning the optimal control of each single player as a function of the population distribution over time and space. The second PDE is the advection equation which is parametrized in the optimal control $u(\cdot)$ and is to be solved in the variable density $m(\cdot)$. It represents how the population behavior (captured by the density) evolves with time if all players are rational and adopt the optimal control $u(\cdot)$. The aim is then to arrive at a fixed point, that is, after making an assumption on $m(\cdot)$, solving the HJB and obtaining a $u(\cdot)$ which, once substituted into the advection equation gives exactly the same density $m(\cdot)$ which we had hypothesized at the beginning of the computation step. When this occurs, such a solution is called *mean–field equilibrium* and represents the asymptotic (when the number of players grows to infinity) solution of a Nash equilibrium.

Let us now look at how we can obtain the HJB equation. This is essentially done in three steps. First, we apply the Bellman Principle which we discussed in the chapter on differential games and dynamic programming, second we use a Taylor expansion and finally we impose a gradient equal to zero as it is usually done when minimizing convex function. In particular, from the Bellman principle we know that the value function $J(x_0, t_0)$ which represents today's cost, can be decomposed as the sum of a stage cost $\min_u[\frac{1}{2}|u_0|^2 + g(x_0, m_0)]$ and a future cost-to-go $J(x_0 + dx, t_0 + dt)$ depending on the future state $x_0 + dx$

which we reach by applying the optimal u. In a nutshell we have:

$$\underbrace{J(x_0,t_0)}_{\text{today's cost}} \quad = \quad \min_u \underbrace{[\frac{1}{2}|u_0|^2 + g(x_0,m_0)]dt}_{\text{stage cost}} + \underbrace{J(x_0+dx,t_0+dt)}_{\text{future cost}}.$$

From expanding the future cost according to Taylor we obtain that

$$J(x_0+dx,t_0+dt) = J(x_0,t_0) + \partial_t J dt + \nabla_x J \dot{x} dt.$$

After computing for the Hamiltonian we have

$$\min_u \underbrace{[\frac{1}{2}|u|^2 + g(x,m) + \partial_t J + \nabla_x J \overbrace{\dot{x}}^{u}]}_{\text{Hamiltonian}} = 0,$$

where we have dropped the index 0. We note that the optimal control is given by $u = -\nabla_x J$ which in turn yields

$$-\partial_t J + \frac{1}{2}|\nabla_x J|^2 = g(x,m) \quad \text{HJB.}$$

12.2 Second-order mean–field game

So far, for sake of simplicity, we considered a simple deterministic dynamics governed by a first-order differential equation. However, the classical mean–field game comes with a stochastic disturbance. More specifically, the underlying dynamics is usually a stochastic dynamics driven by a Brownian motion, which is given by

$$dx = udt + \sigma dB_t,$$

where dB_t is the infinitesimal Brownian motion.

The resulting mean–field game maintains the same structure as in the deterministic case, but now both PDEs involve the second-order derivatives of the value function $v(\cdot)$ and the density $m(\cdot)$ as shown next:

$$\begin{tikzcd}
-\partial_t J + \frac{1}{2}|\nabla_x J|^2 - \frac{\sigma^2}{2}\Delta J = g(x,m) \arrow[d, bend right, "u"'] & \text{(HJB)-backward} \\
\partial_t m + div(m \cdot u(x)) - \frac{\sigma^2}{2}\Delta m = 0 \arrow[u, bend right, "m"'] & \text{(Kolmogorov)-forward}
\end{tikzcd}$$

where $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ denotes the Laplacian operator. Because of the presence of the second derivatives, the above game is called second-order mean–field game. The advection equation is now renamed *Kolmogorov-Fokker-Planck (KFP)* equation. Such an equation is largely used to describe diffusion processes and as such is among the foundations of statistical mechanics and the theory of chaos.

12.3 Infinite horizon: average and discounted cost

Mean-field games admit also a finite horizon formulation which may come in two different forms depending whether the players are patient (farsighted) or impatient (shortsighted). In the former case the formulation involves a long run average cost functional of the form

$$J = E \limsup_{T \to \infty} \frac{1}{T} \int_0^T \Big[\frac{1}{2}|u(t)|^2 + g(x(t), m(\cdot, t))\Big] dt.$$

Then the mean–field game requires solving in $\mathbb{R}^n$ the system

$$\begin{tikzcd}
\bar{\lambda} + \frac{1}{2}|\nabla_x \bar{J}|^2 - \frac{\sigma^2}{2}\Delta \bar{J} = g(x, \bar{m}) \arrow[d, bend right, "u"'] & \text{(HJB)} \\
div(\bar{m} \cdot u(x)) - \frac{\sigma^2}{2}\Delta \bar{m} = 0 \arrow[u, bend right, "\bar{m}"'] & \text{(Kolmogorov)}
\end{tikzcd}$$

In the above formulation, instantaneous cost fluctuations do not count, but what matters is only the long term average cost. Again the structure of the problem is not different from what we have seen so far, though the two equations involve new actors such as the average stage cost $\bar{\lambda}$, the long run average value function $\overline{J}(\cdot)$, and the long run average density $\overline{m}(\cdot)$.

When the players are farsighted the cost involves a discount factor as shown next

$$J = E \int_0^\infty e^{-\rho t} \Big[\frac{1}{2}|u(t)|^2 + g(x(t), m(\cdot, t))\Big] dt.$$

Again the mean–field game requires solving in $\mathbb{R}^n \times [0,T]$ the system

$$
\begin{array}{cl}
-\partial_t J + \frac{1}{2}|\nabla_x J|^2 - \frac{\sigma^2}{2}\Delta J + \rho J = g(x,m) & \qquad \text{(HJB)} \\
u \Big\downarrow \qquad\qquad\qquad\qquad\qquad\qquad \Big\uparrow m & \\
\partial_t m + div(m \cdot u(x)) - \frac{\sigma^2}{2}\Delta m = 0 & \qquad \text{(Kolmogorov)}
\end{array}
$$

12.3.1 Existence and uniqueness

The seminal paper by Lasry and Lions [Lasry and Lions, 2007] also provides general results on existence and uniqueness of mean–field equilibria. These results have also been collected in some lecture notes taken by Pierre Cardaliaguet during a Lions course at the College de France [Cardaliaguet, 2012]. *Existence* is in general proved under the following assumptions (see Theorem 3.1 in [Cardaliaguet, 2012]):

- the running and the terminal cost are uniformly bounded in the space of states and distribution
- the running and the terminal cost are Lypschitz continuous in the space of states and distribution
- the initial probability measure is absolute continuous with respect to the Lebesgue measure.

The above conditions guarantee a certain regularity of the solution in terms of value function and distribution. The third condition in particular excludes the concentration of masses in specific points as the distribution cannot have Dirac impulses. Note that the above conditions guarantee the existence of a *classical solution*. Existence of solutions in a weak sense for very general problems is still an open issue and provides very interesting challenges.

Uniqueness, is in general guaranteed under some monotonicity conditions (see Theorem 3.6 in [Cardaliaguet, 2012]): for the running cost it must hold

$$\int_{\mathbb{R}^d} \Big(g(x,m_1) - g(x,m_2)\Big) d(m_1 - m_2)(x) > 0, \quad \forall m_1, m_2 \in \mathcal{P}, \quad m_1 \neq m_2$$

and for the terminal penalty we must have

$$\int_{\mathbb{R}^d} \Big(G(x,m_1) - G(x,m_2)\Big) d(m_1 - m_2)(x) > 0, \quad \forall m_1, m_2 \in \mathcal{P},$$

where $\mathcal{P}$ is the space of probability distributions. Essentially the above conditions capture scenarios where the higher the concentration of particles around a given point the higher the cost. This phenomenon is generally referred to as *crowd-aversion* and is particularly useful in transportation problems [Lachapelle and Wolfram, 2011].

Explicit computation of equilibrium strategies is possible only in few cases, among which the linear-quadratic case. A successful treatment of linear-quadratic mean–field games is provided by Bardi in [Bardi, 2012].

12.4 Examples

This section presents a few examples borrowed from [Gueant et al., 2010]. These examples are simple and intended to illustrate the generality of the theory and its versatility. The examples span from social science to economics and production engineering.

Example 12.1. (Mexican wave) This is a stylized model capturing mimicry and emulation. The state is $x = [y, z]$, where $y \in [0, L)$ is the horizontal coordinate, and z is the vertical position henceforth referred to as *posture*. Given a continuum of players disposed in the interval $[0, L)$, while the horizontal coordinate of every player is fixed, the posture varies with time and can take any value between 0 and 1, namely

$$z = \begin{cases} 1 & \text{standing} \\ 0 & \text{seated} \end{cases}, \quad z \in (0,1) \quad \text{intermediate.}$$

The dynamics for the posture depends on the input u chosen by the players, which determines the rate of change for the posture as expressed by the first-order law:

$$dz = udt.$$

Essentially, u is the control that every player needs to select. To represent a classical mexican wave scenario, it suffices to consider a penalty on state and distribution of the form:

$$g(x,m) = \underbrace{Kz^{\alpha}(1-z)^{\beta}}_{\text{comfort}} + \underbrace{\frac{1}{\epsilon^2}\int (z-\tilde{z})^2 m(\tilde{y};t,\tilde{z})\frac{1}{\epsilon}s(\frac{y-\tilde{y}}{\epsilon})d\tilde{z}d\tilde{y}}_{\text{mimicry}}.$$

The above term involves two conflicting terms; the comfort of the players which is maximal in the two extreme postures (seated or standing). Indeed the first cost term $z^\alpha(1-z)^\beta$ is concave with minima in zero and one. The second contribution is the mimicry term which penalizes the square deviation $(z-\tilde{z})^2$ of the players posture from its neighbors' posture. The penalty term is weighted by the distance between neighbors (the term $\frac{1}{\epsilon}s(\frac{y-\tilde{y}}{\epsilon})$ represents the Guassian kernel which means that "far neighbors" are less influential than "close neighbors") and the probability $m(\tilde{y};t,\tilde{z})$ that a given neighbor is in a specific posture at a given time.

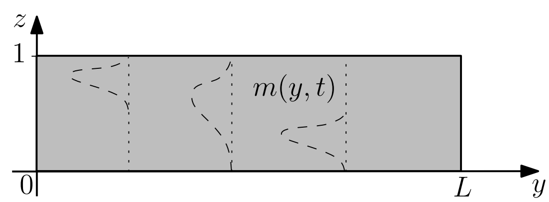

Figure 12.2: Mexican wave: probability that player in position y take on posture z.

Example 12.2. (Meeting starting time) This simple model is in-

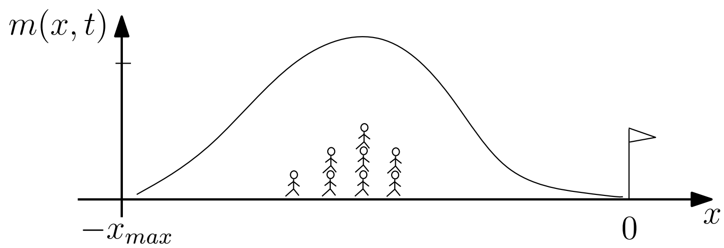

Figure 12.3: Coordination under externality: the meeting starting time example.

tended to capture coordination among players in the presence of an externality (exogenous signal). The story tells that there is a meeting scheduled to start at time t_s in the meeting room based in the origin of the horizontal axis in Fig. 12.3. A continuum of players are initially located in the negative axis and each player needs to select a proper velocity with which to approach the meeting room depending on its expectation of how late the meeting will start given that it will start when at least θ percent of the attendants (θ is the quorum) will have reached the room. In other words, the optimal velocity u depends on every player prediction on how punctual the others are going to be. The underlying dynamics for player i is then

$$dx_i = u_i dt + \sigma dB_t,$$

where the Brownian motion represents stochastic disturbances in approaching the meeting room. Let us denote by $\tilde{\tau}_i = min_s(x_i(s) = 0)$ the arrival time of player i, and by $\bar{t}$ the actual starting time of the meeting.

To complete the model we need to consider a penalty on final state and distribution of type

$$G(x(\tilde{\tau}_i), m(\cdot, \tau_i)) = \underbrace{c_1[\tilde{\tau}_i - t_s]_+}_{\text{reputation}} + \underbrace{c_2[\tilde{\tau}_i - \bar{t}]_+}_{\text{inconvenience}} + \underbrace{c_3[\bar{t} - \tilde{\tau}_i]_+}_{\text{waiting}}.$$

In the above cost we have three different contributions: the first one is a reputation cost which is payed for arriving to late with respect to the scheduled time; the second term is an inconvenience cost associated with arriving too late with respect to the actual starting time; finally the third term is the cost of waiting when one arrives too early with respect to the actual starting time. Given the above model, it is possible to calculate the number of people arrived up to time s, which is given by:

$$F(s) = -\int_0^s \partial_x m(0, v) dv.$$

Then the resulting starting time will be obtained from the inverse function, i.e.,

$$\bar{t} = F^{-1}(\theta).$$

Example 12.3. (Herd behavior) This third example is useful to capture herd behavior in social science. Let x_i be the characteristic or behavior of player i (its political opinion, social behavior, or innovation openness). Assume that the behaviour dynamics is given by

$$dx_i = u_i dt + \sigma dB_t.$$

To represent the typical herd behavior it suffices to consider a penalty of type

$$g(x,m) = \beta(x - \underbrace{\int ym(y,t)dy}_{\text{average}})^2.$$

The above penalty considers the square deviation of the behavior of the player from the average behavior computed over the whole population. The problem can then be formulated as a discounted mean–field game as we have already seen earlier, i.e.,

$$J = E\int_0^{\infty} e^{-\rho t}\Big[\frac{1}{2}|u(t)|^2 + g(x(t), m(\cdot,t))\Big]dt.$$

The resulting mean–field game is then

$$-\partial_t J + \tfrac{1}{2}|\nabla_x J|^2 - \tfrac{\sigma^2}{2}\Delta J + \rho J = g(x,m) \qquad \text{(HJB)}$$

$$\partial_t m + div(m \cdot u(x)) - \tfrac{\sigma^2}{2}\Delta m = 0 \qquad \text{(Kolmogorov)}$$

Example 12.4. (Oil production) Consider a continuum of oil producers each one equipped with a given initial reserve or stock of raw material. A largely adopted stock market model adopts a geometric Brownian motion stochastic process of the form

$$dx = [\alpha_t x + \beta_t u]\,dt + \sigma_t x d\mathcal{B}_t,$$

where $\beta_t u$ is the produced quantity. The penalty (- total income + production costs) appears as

$$g(x,u,m) = -h(\bar{m})u + [\frac{a}{2}u^2 + bu],$$

where $h(\bar{m})$ is sale price of oil, which is decreasing on $\bar{m}$, and $[\frac{a}{2}u^2 + bu]$ are quadratic and linear production costs. The penalty on final state accounts for unexploited reserve:

$$G(x(T)) = \phi|x(T)|^2, \quad \phi > 0.$$

12.5 Robustness

Robustness is here related to the presence of a deterministic adversarial disturbance in addition to the classical stochastic disturbance given by the Brownian motion. "Adversarial" means that of all possible realizations, we will consider the worst-case one, in the same spirit of H_∞ robust optimal control. We illustrate this on the simple example of oil production introduced earlier.

Example 12.5. (Oil production cont'd) For a continuum of oil producers, each one equipped with a given initial reserve or stock of raw material consider the geometric Brownian motion stochastic process

$$dx = [\alpha_t x + \beta_t u + \sigma_t \zeta]\, dt + \sigma_t x d\mathcal{B}_t.$$

The above model describes the time evolution of the reserve. The new term $\sigma_t \zeta$ represents taxation or inflation on the production.

The penalty (- total income + production costs) appears as

$$g(x, u, m, \zeta) = -h(\bar{m}, \zeta)u + [\frac{a}{2}u^2 + bu],$$

where now the sale price of oil $h(\bar{m}, \zeta)$ depends on the disturbance ζ. The idea is to tackle the problem considering the worst-case disturbance as in the book by Başar and Bernhard [Başar and Bernhard, 1995]. This leads to the following inf-sup optimization:

$$\inf_{\{u\}_t} \sup_{\{\zeta\}_t} \mathbb{E}\Big(G(x(T)) + \int_0^T g(x, u, m, \zeta)dt - \gamma^2 \int_0^T |\zeta|^2 dt\Big).$$

Essentially, we look for the infimum with respect to the control u and the supremum with respect to the disturbance ζ. A crucial aspect is the selection of an opportune value for γ which makes the problem not ill-posed.

12.5.1 Worst-case disturbance feedback mean–field equilibrium

The considered set-up leads to a new equilibrium concept, called *worst-case disturbance feedback mean–field equilibrium*, which combines two existing concepts. The first one is the worst-case disturbance feedback

NE derived in the H_∞ literature [Başar and Olsder, 1999], while the second one is the mean–field equilibrium. Note that the worst-case disturbance feedback NE accounts for adversarial disturbances but in the case of a finite number of players. On the contrary, the mean–field equilibrium involves an infinite number of players but in absence of adversarial disturbances. The worst-case disturbance feedback mean–field equilibrium combines both elements, an adversarial disturbance and infinite number of players.

As for the mean–field equilibrium, also the worst-case disturbance feedback mean–field equilibrium requires the solution of the two coupled PDEs displayed below. The first block includes the Hamilton-Jacobi-Isaacs equation (HJI), which returns the value function $J(\cdot)$ and with it also the optimal control $u^*(\cdot)$ and the worst-case disturbance $\zeta^*(\cdot)$. Both control and disturbance are then substituted into the Fokker-Planck-Kolmogorov equation as they both concur in defining the vector field from which we obtain the new density $m(\cdot)$. Again the worst-case disturbance feedback mean–field equilibrium is the fixed point of such a procedure.

$$\boxed{\begin{array}{c} \partial_t J + H_t(x, \partial_x J, m) + \left(\dfrac{\sigma_t}{2\gamma}\right)^2 |\partial_x J|^2 + \dfrac{1}{2}\sigma_t^2 x^2 \partial_{xx}^2 J = 0, \\ \zeta^* = \dfrac{\sigma_t}{2\gamma^2}\partial_x J, \quad u^* = \dfrac{1}{\beta_t}\left[\partial_p H_t(x_t, \dfrac{2\gamma^2}{\sigma_t}\zeta_t^*, m) - \alpha_t x_t\right] \end{array}}$$

$$\big\downarrow {\scriptstyle J,u^*,\zeta^*} \qquad\qquad\qquad\qquad \big\uparrow {\scriptstyle m}$$

$$\boxed{\partial_t m + \partial_x\left(m \partial_p H_t(x, \partial_x J, m)\right) + \frac{\sigma_t^2}{2\gamma^2}\partial_x(m\partial_x J) - \frac{1}{2}\sigma_t^2 \partial_{xx}^2\left[x^2 m\right] = 0}$$

12.6 Conclusions

Mean field games require solving coupled partial differential equations, the HJB equation and the Kolmogorov equation. This chapter has described how robustness can be brought within the picture thus leading to the solution of the HJB under the worst-case disturbance. We have called such a new set-up *robust mean–field games* and the corresponding equilibrium solution as *worst-case disturbance feedback mean–field equilibrium.*

Acknowledgements

These lecture notes have been prepared as a supplement to the series of 12 lectures on game theory to be delivered by the author during a sabbatical period spent as academic visitor at the Department of Engineering Science of the University of Oxford, UK, June-July, 2013, and at the Department of Electrical and Electronic Engineering of the Imperial College London, UK, October-December 2013. The author would like to thank Prof. Antonis Papachristodoulou who promoted and organized the course in Oxford, and Prof. Alessandro Astolfi and Dr. Thulasi Mylvaganam who encouraged and planned the course at Imperial College London. Many thanks are due to the several PhD students, postdocs, and fellows who have attended the courses and have contributed to the improvement of the material with their comments and questions. The author is also sincerely grateful to the reviewers for their careful reading and constructive comments.

References

Y. Achdou, F. Camilli, and I. Capuzzo Dolcetta. Mean field games: numerical methods for the planning problem. *SIAM Journal of Control and Optimization*, 50:77–109, 2012.

E. Altman. Applications of dynamic games in queues. *Advances in Dynamic Games, Annals of the International Society of Dynamic Games*, 7:309–342, 2005.

R. Amir. Continuous stochastic games of capital accumulation with convex transitions. *Games & Economic Behaviors*, 15:111–131, 1996.

R. J. Aumann. Markets with a continuum of players. *Econometrica*, 32(1-2): 39–50, 1964.

R. J. Aumann. Game theory. In *The New Palgrave*, volume 2, London: Macmillan, 1987. J. Eatwell, M. Milgate, and P. Newman, editors.

T. Başar and P. Bernhard. *H^∞ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach.* Birkhäuser, Boston, MA, second edition, 1995.

T. Başar and G.J. Olsder. *Dynamic Noncooperative Game Theory.* Classics in Applied Mathematics. SIAM, Philadelphia, second edition, 1999.

F. Bagagiolo and D. Bauso. Mean-Field Games and Dynamic Demand Management in Power Grids. *Dynamic Games and Applications*, 4(2):155–176, 2014.

M. Bardi. Explicit solutions of some Linear-Quadratic Mean Field Games. *Network and Heterogeneous Media*, 7:243–261, 2012.

D. Bauso and A. Nedić. Dynamic Coalitional TU Games: Distributed Bargaining among Players' Neighbors. *IEEE Transactions on Automatic Control*, 58(6):1363–1376, 2013.

D. Bauso and R. Pesenti. Team theory and person-by-person optimization with binary decisions. *Siam Journal on Control and Optimization*, 50(5): 3011–3028, 2012.

D. Bauso and R. Pesenti. Mean Field Linear Quadratic Games with Set Up Costs. *Dynamic Games and Applications*, 3(1):89–104, 2013.

D. Bauso and J. Timmer. Robust dynamic cooperative games. *International Journal of Game Theory*, 38(1):23–36, 2009.

D. Bauso and J. Timmer. On robustness and dynamics in (un)balanced coalitional games. *Automatica*, 48(10):2592–2596, 2012.

D. Bauso, F. Blanchini, and R. Pesenti. Robust control strategies for multi inventory systems with average flow constraints. *Automatica*, 42:1255–1266, 2006.

D. Bauso, L. Giarré, and R. Pesenti. Consensus in noncooperative dynamic games: a multi-retailer inventory application. *IEEE Transactions on Automatic Control*, 53(4):998–1003, 2008.

D. Bauso, L. Giarré, and R. Pesenti. Distributed consensus in noncooperative inventory games. *European Journal of Operational Research*, 192(3):866–878, 2009.

D. Bauso, F. Blanchini, and R. Pesenti. Optimization of long run average-flow cost in networks with time-varying unknown demand. *IEEE Transactions on Automatic Control*, 55(1):20–31, 2010.

D. Bauso, E. Solan, E. Lehrer, and X. Venel. Attainability in Repeated Games with Vector Payoffs. *INFORMS Mathematics of Operations Research*, Available at http://arxiv.org/abs/1201.6054, 2014.

R. E. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, republished 2003: dover edition, 1957.

D. P. Bertsekas and I. B. Rhodes. On the minimax reachability of target set and target tubes. *Automatica*, 7:233–247, 1971.

J. Bewersdorff. *Luck, logic, and white lies: the mathematics of games.* A K Peters/CRC Press ISBN-13: 978-1568812106, 2004.

D. Blackwell. An Analog of the MinMax Theorem for Vector Payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

O. N. Bondareva. Some Applications of Linear Programming Methods to the Theory of Cooperative Game. *Problemi Kibernetiki*, 10:119–139, 1963.

A. Bressan. Noncooperative Differential Games. A tutorial 2010. available online at http://www.math.psu.edu/bressan/PSPDF/game-lnew.pdf, 2010.

P. Cardaliaguet. Notes on Mean Field Games. P.-L. Lions' lectures, Collège de France, available online at https://www.ceremade.dauphine.fr/∼cardalia/MFG100629.pdf, 2012.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.

V. Chari and P. Kehoe. Sustainable plans. *Journal of Political Economics*, 98:783–802, 1990.

A. Di Mare and V. Latora. Opinion formation models based on game theory. *International Journal of Modern Physics C, Computational Physics and Physical Computation*, 18(9), 2007.

P. Dutta and R. K. Sundaram. The tragedy of the commons? *Economic Theory*, 3:413–426, 1993.

J. Engwerda. *LQ Dynamic Optimization and Differential Games*. John Wiley & Sons, 2005.

J. A. Filar and K. Vrieze. *Competitive Markov decision processes*. Springer, 1996.

A. M. Fink. Equilibrium in a stochastic n-person game. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)*, 28:89–93, 1964.

D. Fudenberg and D. K. Levine. *The theory of learning in games*. MIT press, 1998.

R. Gibbons. *Game Theory for Applied Economists*. Princeton University Press, 1992.

O. Gueant, J. M. Lasry, and P. L. Lions. Mean field games and applications. In *Paris-Princeton Lectures*, pages 1–66. Springer, 2010.

T. Hamilton and R. Mesic. A Simple Game-Theoretic Approach to Suppression of Enemy Defenses and Other Time Critical Target Analyses. RAND Project Air Force, available online at http://www.rand.org/pubs/documented_briefings/DB385.html, 2004.

J. C. Harsanyi and R. Selten. *A General Theory of Equilibrium Selection in Games*. MIT Press, 1988.

S. Hart. Shapley Value. In *The New Palgrave: Game Theory*, pages 210–216, Norton, 1989. J. Eatwell, M. Milgate, and P. Newman, editors.

S. Hart. Adaptive heuristics. *Econometrica*, 73:1401–1430, 2005.

S. Hart and A. Mas-Colell. A General Class of Adaptive Strategies. *Journal of Economic Theory*, 98:26–54, 2001.

S. Hart and A. Mas-Colell. Regret-based continuous-time dynamics. *Games and Economic Behavior*, 45:375–394, 2003.

F. S. Hillier and G. J. Lieberman. *Introduction to Operations Research*. McGraw-Hill, seventh edition, 2001.

Y.-C. Ho. Team decision theory and information structures. *Proceedings IEEE*, 68:644–654, 1980.

M. Y. Huang, P. E. Caines, and R. P. Malhamé. Individual and Mass Behaviour in Large Population Stochastic Wireless Power Control Problems: Centralized and Nash Equilibrium Solutions. In *IEEE Conference on Decision and Control*, pages 98–103, HI, USA, 2003.

M. Y. Huang, P. E. Caines, and R. P. Malhamé. Large Population Stochastic Dynamic Games: Closed Loop Kean-Vlasov Systems and the Nash Certainty Equivalence Principle. *Communications in Information and Systems*, 6(3):221–252, 2006.

M. Y. Huang, P. E. Caines, and R. P. Malhamé. Large population cost-coupled LQG problems with non-uniform agents: individual-mass behaviour and decentralized ϵ-nash equilibria. *IEEE Trans. on Automatic Control*, 52(9): 1560–1571, 2007.

R. Isaacs. *Differential Games: A mathematical theory with applications to warfare and pursuit, control and optimization*. Wiley, (reprinted by dover in 1999) edition, 1965.

J. S. Jordan. Three problems in learning mixed-strategy nash equilibria. *Games and Economic Behavior*, 5:368–386, 1993.

B. Jovanovic and R. W. Rosenthal. Anonymous sequential games. *Journal of Mathematical Economics*, 17:77–87, 1988.

M. Kandori, G. J. Mailath, and R. Rob. Learning, Mutation, and Long-run Equilibria in Games. *Econometrica*, 61:29–56, 1993.

A. Lachapelle and M.-T. Wolfram. On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation Research Part B*, 15(10):1572–1589, 2011.

A. Lachapelle, J. Salomon, and G. Turinici. Computation of Mean Field Equilibria in Economics. *Math. Models Meth. Appl. Sci.*, 20:1–22, 2010.

J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. I Le cas stationnaire. *Comptes Rendus Mathematique*, 343(9):619–625, 2006a.

J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. II Horizon fini et controle optimal. *Comptes Rendus Mathematique*, 343(10):679–684, 2006b.

J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2:229–260, 2007.

E. Lehrer. Approachability in infinite dimensional spaces. *International Journal of Game Theory*, 31(2):253–268, 2002.

E. Lehrer and S. Sorin. Minmax via differential inclusion. *Convex Analysis*, 14(2):271–273, 2007.

E. Lehrer, E. Solan, and D. Bauso. Repeated Games over Networks with Vector Payoffs: the Notion of Attainability. In *Proceedings of Intl. Conference on Network Games, Control and Optimization (NETGCOOP 2011)*, pages 1–5, 2011.

D. Levhari and L. Mirman. The great fish war: An example using a dynamic Cournot-Nash solution. *The Bell Journal of Economics*, 11(1):322–334, 1980.

D. Liberzon. *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton University Press, 2012.

J. Marschak and R. Radner. *Economic Theory of Teams*. Yale University Press, New Haven, CT, USA, 1972.

M. Maschler, E. Solan, and S. Zamir. *Game Theory*. Cambridge University Press, 2013.

J. F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–66, 1981.

J. D. Morrow. *Game Theory for Political Scientists*. Princeton, NJ: Princeton University Press, 1994.

J. F. Nash Jr. Equilibrium points in n-person games. *Proc. National Academy of Sciences*, 36(1):48–49, 1950.

J. F. Nash Jr. Non-cooperative games. *Annals of Math.*, 54(2):286–295, 1951.

A. Neyman and S. Sorin. Stochastic games and applications. In *NATO Science Series*. Kluwer, 2003.

N. Noam, T. Roughgarden, E. Tardos, and V. Vazirani. *Algorithmic Game Theory*. Cambridge, UK: Cambridge University Press ISBN 0-521-87282-0, 2007.

A. S. Nowak. On a new class of nonzero-sum discounted stochastic games having stationary nash equilibrium points. *International Journal of Game Theory*, 32:121–132, 2003.

M. J. Osborne and A. Rubinstein. *Bargaining and Markets*. Series in economic theory, econometrics, and mathematical economics, Academic Press, 1990.

M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT press, Cambridge, MA, 1994.

A. Ozdaglar. Game theory with engineering applications. MITOPENCOURSEWARE, available at http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-254-game-theory-with-engineering-applications-spring-2010/, 2010.

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

B. Polak. Game theory. Open Yale Course available at http://oyc.yale.edu/economics/econ-159, 2007.

P. V. Reddy and J. C. Engwerda. Pareto optimality in infinite horizon linear quadratic differential games. *Automatica*, 49(6):1705–1714, 2013.

W. Saad, Z. Han, M. Debbah, A. Hjörungnes, and T. Başar. Coalitional game theory for communication networks: a tutorial. *IEEE Signal Processing Magazine, Special Issue on Game Theory*, 26(5):77–97, 2009.

Y. E. Sagduyu and A. Ephremides. Power control and rate adaptation as stochastic games for random access. In *Proc 42nd IEEE Conference on Decision and Control*, volume 4, pages 4202–4207, 2003.

W. H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT press, 2010.

L. S. Shapley. Stochastic games. In *Proc Nat Acad Sci USA*, volume 39, pages 1095–1100, 1953.

L. S. Shapley. Some topics in two-person games. *Ann. Math. Studies*, 5:1–8, 1964.

L. S. Shapley. On balanced sets and cores. *Naval Research Logistics Quarterly*, 14:453–460, 1967.

Y. Shoham and K. Leyton-Brown. *Multiagent Systems Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

J. M. Smith. Game theory and the evolution of fighting. In *On Evolution*. Edinburgh University Press, 1972.

J. M. Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

J. M. Smith and G. R. Price. The Logic of Animal Conflict. *Nature*, 246: 15–18, 1973.

E. Solan. Stochastic games. In *Encyclopedia of Database Systems*. Springer. Also available at http://www.math.tau.ac.il/%7Eeilons/encyclopedia.pdf, 2009.

A. S. Soulaimani, M. Quincampoix, and S. Sorin. Approchability theory, discriminating domain and differential games. *SIAM Journal of Control and Optimization*, 48(4):2461–2479, 2009.

S. Tijs. *Introduction to Game Theory*. Hindustan Book Agency, 2003.

N. Vieille. Equilibrium in 2-person stochastic games I: A reduction. *Israel Journal of Mathematics*, 119:55–91, 2000.

J. von Neumann. Zur theorie der gesellschaftspiele. *Mathematische Annalen*, 100:295–320, 1928.

J. von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1944.

H. von Stackelberg. *Marktform und Gleichgewicht*. Springer Verlag, (An English translation appeared in 1952 entitled *The Theory of the Market Economy*, published by Oxford University Press, Oxford, England.), Vienna, 1934.

J. Weibull. *Evolutionary game theory*. MIT press, 1995.

G. Y. Weintraub, C. Benkard, and B. Van Roy. Oblivious Equilibrium: A Mean Field Approximation for Large-Scale Dynamic Games. *Advances in Neural Information Processing Systems, MIT press*, 2005.

D. W. K. Yeung and L. A. Petrosjan. *Cooperative stochastic differential games*. Springer series in Operations Research and Financial Engineering, 2006.

H. P. Young. The Evolution of Conventions. *Econometrica*, 61:57–84, 1993.

Q. Zhu and T. Başar. A multi-resolution large population game framework for smart grid demand response management. In *Proceedings of Intl. Conference on Network Games, Control and Optimization (NETGCOOP 2011)*, Paris, France, 2011.