An Introduction to Learning in Games

Lillian Ratliff

March 6, 2021

Contents

Preface

This set of lecture notes is a **work in progress** being developed by myself, Lillian Ratliff, and Shankar Sastry for courses we each teach that either focus on game theory or have subcomponents on the matter. Part III of these lecture notes is forthcoming, as are key parts of the appendix which review preliminary material.

Chapter 1

Introduction

Game theory deals with situations of conflict between players or agents. Each participating player can partially control the environment or situation, but no player has full control. Moreover, each player has certain personal preferences over the set of possible outcomes and strives to obtain that outcome which is most profitable to them.

> *We wish to find the mathematically complete principles which define "rational behavior" for the participants in a social economy, and to derive from them the general characteristics of that behavior. And while the principles ought to be perfectly general—i.e., valid in all situations—we may be satisfied if we can find solutions, for the moment, only in some characteristic special cases.* (von Neumann and Morgenstern '47)

Game theory is a *normative theory* in the sense that it aims to prescribe what each player in a game should do in order to promote their interests optimally—i.e., which strategy each player should play such that their partial influence on the situation benefits them the most. The aim of game theory, as a mathematical theory, is to provide a solution—i.e., characterization of *rational behavior*—for every game.

Towards understanding how players arrive at an equilibrium, one natural perspective is that they grapple with one another in a tâtonnement process such as following a *learning algorithm*. The goal is often to characterize the expected outcomes of interactions between rational agents implementing some *natural* learning rule. A recent view in game theory is that learning in games provides *axiomatic backing* for equilibrium in the sense that they arise and can be explained through the outcomes of iterative competitions for optimality. This perspective motivates the analysis of learning algorithms which reflect any natural structure present in a problem and doing so often suggests refinements of standard equilibrium notions.

Furthermore, in machine learning there has been an emergence of game-theoretic constraints—e.g., interactions with other agents, adversarial environments, or market structures—in systems where learning algorithms are deployed and a rise in the number of machine learning problems that are being formulated as games such as generative adversarial networks. Indeed, learning algorithms are now often embedded in real-world systems and tasked with acquiring information to enable effective decision-making. However, learning agents are rarely acting in isolation within complex systems; instead they are typically in the presence of multiple autonomous agents who may be optimizing conflicting objectives and hence competing. Simply put, game theoretic constraints are a naturally occurring phenomena in the real world.

This fundamentally changes what outcomes can be expected of algorithms designed for static environments and the types of algorithms that can achieve desired objectives. The emergence of competition introduces a number of distinct questions and challenges in the study of learning and decision-making. To begin, it requires rethinking what it even means to act optimally or solve a problem. Typically, doing so necessitates considering an equilibrium notion as a solution concept. However, even narrowing down a proper notion of equilibrium is a nuanced discussion without a unifying resolution. Then, given a solution concept, effective decision-making demands accounting for the inherently dynamic and non-stationary environment. This means it is imperative to not only consider the stationary objective being optimized, but also how interactions with competing agents and the environment affect the pursuit of that objective while learning.

With these motivations in mind, this set of lecture notes is divided into three key parts:

Part I. Normal Form Games

In this part of the lecture notes, we focus on reviewing basic game theoretic constructs. In particular, the focus is on normal form games with either finite actions spaces or continuous action spaces. We review equilibrium concepts, existence and uniqueness and important classes of games such as *convex*, *potential*, *bilinear/bimatrix*, etc. We also provide quintessential examples to help facilitate the discussion and illustrate concepts.

Part II. Learning in Games

Part III. Application to Learning-Based Systems

Part I

Normal Form Games

The focus of Part I is on the review of normal form games on both finite and continuous action spaces. In the chapters contained in Part I, we formalize the definition of a game and equilibrium concepts.

Chapter 2

Introduction to Normal Form Games

Game theory is the study of decision making in the presence of *competition*—that is, when one individual's decisions affect the outcome of a situation for all other individuals involved. We refer to these individuals as players or agents interchangeably. Game theory is a broad discipline that influences and is itself influenced by various other domains of study including operations research, economics, control theory, computer science, psychology, biology, sociology and so on.

A **game** is defined by specifying the *players*, *rules and objectives*, and *information structure*. The players are the agents that make decisions. Often *player* and *agent* are used interchangably. The rules define the actions allowed by the players and their effects, and the objective is the goal of each player. The information structure specifies what each player knows before making each decision.

For a mathematical solution to a game, one further needs to make assumptions on the players' *rationality*. The models we study assume that each decision-maker is *rational* in the sense that they are aware of their alternatives, form expectations about any unknowns, have clear preferences, and choose their action deliberately after some process of optimization. As an example, when we assume players always pursue their best interests to fulfill their objectives, do not form coalitions, and do not trust each other, we refer to such games as being *non-cooperative*. More specifically, a **noncooperative game** is a game in which there are no possibilities for communication, correlation or (pre)commitment, except for those that are explicitly allowed by the rules. In this set of lecture notes, we will treat the non-cooperative case.

Studying non-cooperative solutions for games **played by humans** is some what of "the big lie" in game theory. That is, humans are not rational and are often altruistic or cooperative by their nature. None-the-less, this has by no means prevented economists (as well as computer scientists and engineers) from studying them. When pursuing this approach one should not be overly surprised by finding solutions of "questionable ethics."

One of the greatest contributions of non-cooperative game theory is that it allows one to find "problematic" solutions or unintended outcomes to games, and often indicates how to "fix" the games so that these solutions are prevented or mitigated. The latter falls under the heading of **mechanism design**.

On the other hand, plenty of practical scenarios are such that "players" are abstractions for decision processes not affected by human reasoning and as such, we can attempt to find non-cooperative solutions without questioning their validity. For instance, robust design (robust control, safety/security, decision-making under uncertainty, etc.) and evolutionary biology are good examples.

2.1 Mathematical Formalism

There are two categories of games: normal form games (sometimes referred to as strategic form games) and extensive form games. The two types of games are informally described as follows:

- A **normal form game** is a model of interactive decision-making in which each player chooses their plan of action once and for all, and all players' decisions are made simultaneously. That is, when choosing a *strategy* (i.e., a plan of action) each player is not informed of the strategy chosen by any other player.
- An **extensive game** specifies the possible orders of events: each player can consider their strategy not only at the beginning of the game, but also whenever he has to make a decision. Often at each of these decision points, new information is revealed to one or more of the players.

In this set of lecture notes, we focus only on normal form games. It is important to first fully understand normal form games before proceeding to extensive form games, which are much more challenging to analyze with often very non-intuitive outcomes. We leave this to a second course on game theory. Interested students may read further about extensive form games in the books by Osborne and Rubinstein in [42] or by Başar and Olsder [5] for instance.

To formally define a normal form game, we first need the notion of a preference revelation. Denoted $\succsim_i$ for player i, the preference relation provides an ordering on the utility or value of an action to player i relative to another action. For example, if $x, x' \in X$, then $x \succsim_i x'$ if x is strictly preferred to x'.

Definition 2.1. A **normal form game** consists of the following:

a. a finite set $\mathcal{N} = \{1, \ldots, N\}$ which indexes the N players,
b. a nonempty set X_i for each player $i \in \mathcal{N}$ which is referred to as the action space for player i, where $x \in X$ is called an action profile,
c. and a preference relation $\succsim_i$ on $X = \prod_{i \in \mathcal{N}} X_i$ for each player $i \in \mathcal{N}$.

If the set X_i of actions of every player i is finite then the game is finite (cf. Chapter 3) and, on the other hand, if X_i is continuous for each player, the the game is continuous (cf. Chapter 4).

Recalling that we previously stated that a game is defined by specifying the players, rules and objectives, and information structure. In the above definition of a normal form game, we have defined the players $\mathcal{N}$, and the rules $X = \prod_i X_i$ and objectives $\succsim = (\succsim_1, \ldots, \succsim_N)$. We have not stated the information structure, but implicitly normal form games assume that each player knows the preference relation of each other player and their action spaces. In addition, we assume they are rational in the sense previously described at the beginning of the chapter, and that the game is non-cooperative.

The high level of abstraction of this model allows it to be applied to a wide variety of situations. A player may be an individual human being or any other decision-making entity like a government, a board of directors, the leadership of a revolutionary movement, or even a flower or an animal.

Under a many circumstances, the preference relation $\succsim_i$ of player i in a normal form game can be represented by a *payoff* or *utility* function $u_i : X \to \mathbb{R}$, in the sense that, for any $x, x' \in X$,

$$u_i(x) \geq u_i(x') \iff x \succsim_i x'.$$

We refer to values of such a function as payoffs (or utilities). Often we define a game not by a utility function but by a cost function $f_i(x)$ which captures the loss experienced to player i under action profile $x \in X$. In this case, for any $x, x' \in X$,

$$f_i(x) \leq f_i(x') \iff x \succsim_i x'.$$

Frequently we specify a player's preference relation by giving a payoff function that represents it. In such a case we denote the game by

$$\mathcal{G} = (\mathcal{N}, X, (u_i)_{i \in \mathcal{N}})$$

as opposed to the equivalent tuple

$$\mathcal{G} = (\mathcal{N}, X, (\succsim_i)_{i \in \mathcal{N}}).$$

We will use the notation x_{-i} to denote the actions of all player excluding player i—that is,

$$x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \in \prod_{j \neq i} X_j.$$

Moreover, by a slight abuse of notation, we write a joint action profile x as $x = (x_i, x_{-i})$ and denote player i's utility under that profile as $u_i(x_i, x_{-i})$.

Example 2.2 (Prisoner's Dilemma)**.** Consider a game in which there are two members of an organized crime unit indexed by $\mathcal{N} = \{1, 2\}$ that are arrested for a crime. The agents are kept in solitary confinement and cannot communicate. Moreover, the prosecutors do not have sufficient evidence to convict them on the main charge but have enough evidence for a petty crime. The players can choose to stay silent or betray on another (cooperate or defect).

The outcomes are

	P2 silent	P2 betray
P1 silent	(-1,-1)	(-3,0)
P1 betray	(0,-3)	(-2,-2)

The preferences are thus

- If player 2 stays silent, then player 1 prefers to betray since $0 \geq -1$
- If player 2 betrays, then player 1 prefers to betray since $-2 \geq -3$
- If player 1 stays silent, then player 2 prefers to betray since $0 \geq -1$
- If player 1 betrays, then player 2 prefers to betray since $-2 \geq -3$

This means, in terms of utilities

$$u_1(S, S) \leq u_1(B, S) \text{ and } u_2(S, S) \leq u_2(S, B)$$

That is, the profile $x = (S, S)$ is strictly dominated by the action B for any player which in turn means that $x^* = (B, B)$ is an equilibrium.

2.2 Types of Games

In this section and from here forward, we will define concepts from the perspective that players have *cost function* f_i and hence, are cost minimizers. When it is easier or more appropriate to have players as utility maximizers with utility functions u_i, we will make this clear in the written context.

There are several ways to create a *taxonomy* for games. We can distinguish them based on the information available to agents (cf. the previous discussion on extensive form versus normal form games), the rules of the game (e.g., order of play), or the objectives of players. Even in terms of the objectives, there are several ways to distinguish between games: e.g., it could be based on how payoffs are realizes or on the assumptions of the structure or class of cost functions players use.

In the first part of these lecture notes, we will primarily focus on full information games. When we go to the second part on learning in games, we will discuss partial information and different feedback structures on the learning process.

2.2.1 Characterizing Games Based on Rules

There are many ways in which the rules of the game can lead to methods of categorizing and distinguishing classes of games. In this set of lecture notes, we will only discuss one main way which is central to a lot of game theoretic formulations: the order of play.

The game can either be a simultaneous play game in which it is assumed that players make their choices at the same time, or it can be a hierarchical play game in which it is assumed that there is an order of play between the players. For example, if there are two players, one player assumes the role of the *leader*—meaning it plays first—and the other player assumes the role of the *follower*. Games with a hierarchical order of play are often referred to as *Stackelberg games* [52]. In a *Stackelberg game*, the leader and follower aim to solve the following optimization problems, respectively:

$$\min_{x_1 \in X_1} \{ f_1(x_1, x_2) \big| \; x_2 \in \arg\min_{y \in X_2} f_2(x_1, y) \}, \tag{L}$$

$$\min_{x_2 \in X_2} f_2(x_1, x_2). \tag{F}$$

This contrasts with a *simultaneous* play game in which each player i is faced with the optimization problem

$$\min_{x_i \in X_i} f_i(x_i, x_{-i}). \tag{2.1}$$

Stackelberg-type games are keenly important for mechanism design as well as formulations of robust machine learning, adversarial learning, and strategic classification—or more generally, learning from strategically generated data.

There are different, but natural, equilibrium concepts for the type of game based on the order of play. In particular, there are equilibrium concepts which respect the structure of the order of play and hence the game itself. For certain classes of games, we will see there is tight connection between the different equilibrium notions. In Section 2.3 we introduce these concepts and then make some initial observations about the connections between them. Further details on how they are connected are discussed in Chapters 3 and 4, as well as in Part II where learning dynamics play a key role.

2.2.2 Characterizing Games Based on Objectives

Regarding how payoffs are realized, there are several main types of games: zero-sum games, constant-sum games, common-payoff games, and non-zero-sum games (or general-sum games).

Definition 2.3. A game $\mathcal{G} = (\mathcal{N}, X, (f_i)_{i\in\mathcal{N}})$ is constant sum if

$$\sum_{i\in\mathcal{N}} f_i(x) = c, \ \forall\ x \in X, \tag{2.2}$$

for some constant c. If $c = 0$, then the game is zero-sum, and if the game is not constant or zero-sum then it is non-zero-sum or general sum.

Prisoner's dilemma is a type of non-zero sum game, though it gives rise to the non-intuitive outcome that players should choose a strategy that leads to the worst possible outcome even though there is a socially cooperative solution with better payoffs across the board. Another interesting non-zero sum game is given below.

Example 2.4 (Bach vs Stravinsky)**.** General sum games tend to include elements of both coordination and competition. A pair of friends wish to go to the orchestra, and they can select among two options: "Bach" or "Stravinsky". They much prefer to go together rather than to separate showings, but while the player 1 prefers Bach (B), the player 2 prefers Stravinsky (S).

The outcomes are

	P2 Bach	P2 Stravinksy
P1 Bach	(2,1)	(0,0)
P1 Stravinsky	(0,0)	(1,2)

It turns out that coordinating on either (B, B) or (S, S) is an equilibrium—i.e.,

$$u_1(B, B) = 2 \geq u_1(S, B) = 0 \text{ and } u_2(B, B) = 1 \geq u_2(B, S) = 0$$

as can be similarly shown for (S, S). These are not the only two equilibria however. We will see later that randomizing can also be an equilibrium.

Typically, zero-sum games are taken to be between two players.[1] In this case, the game is then defined by a single cost function f such that $\mathcal{G} = (\mathcal{N}, X, (f, -f))$.

Example 2.5 (Matching Pennies)**.** A classical example of a zero-sum game is the game of Matching Pennies. In this game, each of the two players has a penny, and independently chooses to display either heads or tails. The two players then compare their pennies. If they are the same then player 1 pockets both, and otherwise player 2 pockets them. The outcomes are

	P2 heads	P2 tails
P1 heads	(1,-1)	(-1,1)
P1 tails	(-1,1)	(1,-1)

[1]Note that another common game type is network zero-sum games and pairwise zero-sum games. In the former case, the condition in (2.2) defines the zero-sum nature of the network game. The network defines which players compete with one another—i.e., cost function dependencies on player actions. In the latter case, players that compete with one another do so in pairwise zero-sum interactions. In this case, not only does (2.2) hold, but it is also the case that $f_i(x) + f_j(x) = 0$ if in the network there is an edge between player i and player j, indicating that they compete with one another.

We will see that the best thing players can do (such that they do not have an incentive to deviate) is to randomize. That is, play heads and tails with equal probability.

Rock paper scissors is another common zero sum game in which there are three strategies per player such that rock beats scissors, paper beats rock, and scissors beat paper. In this game, it is also an ”equilibrium” strategy to randomize.

Another important type of game is a *common-payoff* game or *pure coordination* games.

Definition 2.6. A common-payoff game is a type of constant sum game such that for all profiles $x \in X$ and any two agents i and j, $u_i(x) = u_j(x)$.

In such games the agents have no conflicting interests; their sole challenge is to coordinate on an action that is maximally beneficial to all.

Example 2.7 (Coordination Game)**.** Imagine two drivers driving towards each other in a country having no traffic rules, and who must independently decide whether to drive on the left or on the right. If the drivers choose the same side (left or right) they get a high utility, and otherwise they get a low utility. The outcomes are

	P2 left	P2 right
P1 left	(1,1)	(0,0)
P1 right	(0,0)	(1,1)

Clearly, players do not have an incentive to deviate from the coordinated action profiles (left,left) and (right,right).

2.3 Equilibrium Concepts for Simultaneous and Hierarchical Play

As noted above, there are very natural equilibrium concepts that respect the order of play in the game.

Simultaneous Play. For simultaneous play games, the quintessential equilibrium notion (and one that is central to this set of notes) is that of a *Nash equilibrium* [41]. For x^* to be a Nash equilibrium it must be that no player i has an action yielding an outcome that they prefer to that generated when x_i^* is played and everyone else plays x_{-i}^*. More formally, a Nash equilibrium is defined as follows.

Definition 2.8. A Nash equilibrium of $\mathcal{G} = (\mathcal{N}, X, \succsim)$ is a profile $x^* = (x_1^*, \ldots, x_N^*) \in X$ of actions with the property that for every player $i \in \mathcal{N}$,

$$(x_i^*, x_{-i}^*) \succsim_i (x_i, x_{-i}^*), \quad \forall\, x_i \in X_i.$$

Analogously, we can define the concept using the cost function representation of preferences.

Definition 2.9. A Nash equilibrium of $\mathcal{G} = (\mathcal{N}, X, (f_i)_{i\in\mathcal{N}})$ is a profile $x^* = (x_1^*, \ldots, x_N^*) \in X$ of actions with the property that for every player $i \in \mathcal{N}$,

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*), \quad \forall\, x_i \in X_i.$$

Another useful concept is the best response set with which we can restate the above definition. For any $x_{-i} \in X_{-i}$ define $B_i(x_{-i})$ to be

$$B_i(x_{-i}) = \{x_i|\ f_i(x_i, x_{-i}) \leq f_i(x_i', x_{-i}),\ \ \forall\ x_i' \in X_i\}.$$

Definition 2.10. A Nash equilibrium is an action profile $x^* = (x_1^*, \ldots, x_N^*) \in X$ for which

$$x_i^* \in B_i(x_{-i}^*),\ \ \forall\ i.$$

An alternative expression which gives us some intuition for refinements of Nash is to write the best response map as follows:

$$x_i^* \in \arg\min_{x_i} f_i(x_i, x_{-i}^*)$$

The action spaces X_i can be finite or infinite—in particular, a continuum.[2] In either case, both definitions above for Nash can be seen as defining a *pure strategy Nash equilibrium* or a *Nash equilibrium in pure strategies*. Furthermore, if the action spaces are finite—say $X_i = \{1, \ldots, m_i\}$ but we consider the probability simplex

$$\Delta(X_i) = \left\{ x_i :\ \sum_{j=1}^{m_i} x_{ij} = 1,\ 0 \leq x_{ij} \leq 1 \right\}$$

then the definitions above capture *mixed strategy Nash equilibria* where each x_{ij}^* denotes the probability that player i will player strategy $j \in X_i$ in equilibrium. This distinction between pure and mixed strategy equilibrium concepts will be further discussed in the subsequent chapters.

Example 2.11 (Bach vs Stravinsky Take 2)**.** Recall the general sum Bach-Stravinsky game in which the outcomes are

	P2 Bach	P2 Stravinksy
P1 Bach	(2,1)	(0,0)
P1 Stravinsky	(0,0)	(1,2)

We saw that coordinating on either (B, B) or (S, S) are both pure Nash equilibria. These are not the only two equilibria however. Now that we have a notion of a mixed Nash we can see this.

Suppose that $x_2 = (p, 1-p) \in \Delta(X_2)$. Then if the player 1, also mixes between their two actions, they must be indifferent (i.e., utilities must be the same) between them, given player 2's strategy. Indeed, otherwise they would be better off switching to a pure strategy according to which they only played the better of their actions (that is if they are not indifferent they would keep putting more mass on the strategy they prefer until its a pure strategy). Then

$$u_1(S, x_2) = 2p + 0(1-p) = u_1(B, x_2) = 0p + 1(1-p) \implies p = \frac{1}{3}$$

Now, since player 2 is mixing, they must also be indifferent between actions:

$$u_2(x_1, B) = q1 + (1-q)0 = u_2(x_1, S) = 0q + (1-q)2 \implies q = \frac{2}{3}$$

So that

$$(x_1^*, x_2^*) = ((2/3, 1/3), (1/3, 2/3))$$

[2]The case of infinite but countable is rarely encountered so we do not really comment on that case in these lecture notes.

Hierarchical Play. In the hierarchical play setting (or Stackelberg game), the most common equilibrium notion is a Stackelberg equilibrium. In contrast to the Nash concept, the roles of players are not symmetric in that one player has the ability to enforce their strategy on the other player(s) which leads to the hierarchical model of play. In this case, the *leader* takes into consideration that the follower will act in accordance to its objective given the decision of the leader—that is, the follower plays a best response. Yet, both players make their decisions simultaneously. In particular, the notion of Stackelberg equilibrium we consider is for single-act games.[3]

This equilibrium notion takes care to define properly because the leading agent's optimization problem is subjected to a constraint which itself is an optimization problem, thereby leading to a bilevel optimization problem for the leader.

Bilevel optimization problem may not be well-behaved when the inner problem leads to a disjoint set and hence, a discontinuous optimization problem for the leader. This being said, under certain regularity conditions (which will be further discussed in Chapters 3 and 3) we can formulate a clear definition of the equilibrium concept. When the inner problem has multiple solutions (e.g., multiple disconnected global minima), in the bilevel optimization literature both *pessimistic* and *optimistic* solutions have been defined in whcih the leader assumes the follower plays either the worst case solution or the best case solution amongst the set of global minima—parameterized by the leader's choice variable—respectively. In this set notes, we will not go into great detail on these two cases. Instead we will focus on situations where either regularity conditions ensure such a situation does not arise or in which a local notion of equilibrium is well-defined.

Consider a two player game for simplicity, $\mathcal{G} = (\mathcal{N}, X, (f_1, f_2))$, where without loss of generality player 1 is the leader and player 2 is the follower. That is, player 1 plays first and then player 2 follows. This means that player 1 should take into consideration that player 2 is going to play a best response to its choice, and should therefore strategize with this in mind. A Stackelberg equilibrium for $\mathcal{G}$ is defined as follows.

Definition 2.12. The strategy $x_1^* \in X_1$ is a Stackelberg solution for the leader if, $\forall x_1 \in X_1$,

$$\sup\nolimits_{x_2 \in \mathcal{R}_{X_2}(x_1^*)} f_1(x_1^*, x_2) \leq \sup\nolimits_{x_2 \in \mathcal{R}_{X_2}(x_1)} f_1(x_1, x_2),$$

where $\mathcal{R}_{X_2}(x_1) = \{y \in X_2 | f_2(x_1, y) \leq f_2(x_1, x_2), \forall x_2 \in X_2\}$ is player 2's reaction set in response to x_1. Moreover, (x_1^*, x_2^*) for any $x_2^* \in \mathcal{R}_{X_2}(x_1^*)$ is a Stackelberg equilibrium on $X_1 \times X_2$.

2.3.1 Additional Equilibrium Notions

There are many refinements of the above equilibrium concepts, both in terms of the degree of coordination allowed between players (and nature) as well as *local* notions.

In Chapter 3, we will discuss some of the refinements related to the former, and in Chapter 4, we will cover the latter.

Correlated Equilibria. For example, a Nash equilibrium in mixed strategies is such that the probability distribution on $X = X_1 \times \cdots \times X_N$ is independent across players. Indeed, let's introduce a little more notation to make this clear. For joint strategy profile $a \in X$ where $a_i \in X_i =$

[3] Often extensive form game representations are used to analyze Stackelberg equilibria in normal-form games with an order of play. We will not cover this analysis in this course but one can read about it in [5].

$\{1,\ldots,m_i\}$ and $a_{-i} \in X_{-i}$, we denote the expected utility given the mixed strategy $x = \prod_{i\in\mathcal{N}} x_i$ by

$$\mathbb{E}_{a\sim x}[u_i(a_i, a_{-i})]$$

Here strategies a_i are sampled from x_i independently across players as the notation $x = \prod_{i\in\mathcal{N}} x_i$ indicates. If on the other hand, players are allowed to coordinate via an external signal or *recommendation*, then the distribution x is no longer independent across players, and this leads to the notion of a correlated equlibirium. This is the first natural refinement of a Nash equilibrium. That is, all Nash equilibria are correlated equilibria but the other direction does not hold:

$$\text{Pure Nash} \subset \text{Mixed Nash} \subset \text{Correlated Equilibria}$$

This is an important refinement since many learning algorithms for mixed strategies in finite games can only be guaranteed to achieve correlated equilibria.

Recall that in a Nash equilibrium, players randomize over strategies independently. This is one of the primary distinctions between mixed Nash equilibria and correlated equilibria: in a correlated equilibrium there is a correlating mechanism (signal) that the agents use as a reference point in determining a utility maximizing strategy. Due to this, the agents are no longer randomizing independently.

In fact, it may be the case for games with multiple Nash equilibria that we want to allow for randomization between Nash equilibria by some form of communication prior to the play of the game. Consider the following example in which the coordinating mechanism is a coin flip (Bernoulli random variable).

Example 2.13 (Bach v. Stravinsky Take 3)**.**

P1\P2	Bach	Stravinsky
Bach	$(1,2)$	$(0,0)$
Stravinsky	$(0,0)$	$(2,1)$

Suppose the players flip a coin and go to Bach if the coin is heads and Stravinksy otherwise.

The payoff to the players in this case is $(1/2(1+2), 1/2(2+1)) = (3/2, 3/2)$ and this is not a Nash equilibrium.

Example 2.14. (**Chicken-Dare Game**.) As another example, think about a traffic light. As people come to an intersection the light suggests to them what to do (obviously, in reality the traffic light is not random in its assignment of red or green, but as a thought experiment it serves its purpose). This is often called the Chiken-Dare game, and the payoffs are as follows:

P1\P2	**D**are	**C**hicken
Dare	$0,0$	$7,2$
Chicken	$2,7$	$6,6$

There are two pure Nash equilibria (D,C) and (C,D)

$$2 = u_2(D,C) \geq u_2(D,D) = 0$$

$$7 = u_1(D,C) \geq u_1(C,C) = 6$$

and one mixed Nash where the probability of $((D,C),(D,C))$ is $((1/3,2/3),(1/3,2/3))$:

$$\begin{bmatrix}\frac{1}{3}\\ \frac{2}{3}\end{bmatrix}^\top \begin{bmatrix}0 & 7\\ 2 & 6\end{bmatrix}\begin{bmatrix}\frac{1}{3}\\ \frac{2}{3}\end{bmatrix} = \begin{bmatrix}\frac{4}{3}\\ \frac{19}{3}\end{bmatrix}^\top \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} = 4.667 = \frac{14}{3} \geq \begin{bmatrix}a\\ b\end{bmatrix}^\top \begin{bmatrix}0 & 7\\ 2 & 6\end{bmatrix}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} = \begin{bmatrix}a\\ b\end{bmatrix}^\top \begin{bmatrix}\frac{14}{3}\\ \frac{14}{3}\end{bmatrix} = \frac{14}{3}(a+b) = \frac{14}{3}$$

Now suppose that prior to playing the game the players performed the following experiment:

- Player 1 chooses ball from bag with three balls labeled C, C, D
- Player 2 chooses ball from bag with remaining two balls
- Player 1 and player 2 play according to the ball they get

It can be verified that there is no incentive to deviate from such an agreement since the suggested strategy is best in expectation.

This experiment is equivalent to having the following strategy profile chosen for the players by some third party, a **correlation device**:

	Dare	**C**hicken
Dare	0	1/3
Chicken	1/3	1/3

Suppose player 1 gets a ball with C. Then player 2 will play C with probability 1/2 and d with probability 1/2. The expected utility of D is $0(1/2) + 7(1/2) = 3.5$ and the expected utility of C out is $2(1/2)+6(1/2) = 4$. So, the player would prefer to Chicken out. On the other hand, suppose player 1 gets ball D, he would not want to deviate supposing the other player played their assigned strategy since he will get 7 (the highest payoff possible).

The expected payoff for this game is $7(1/3) + 2(1/3) + 6(1/3) = 5$ which is greater than the Nash payoff 14/3.

This equilibrium (using the correlation device) is a *correlated equilibrium* since the players play is correlated.

Local notions of Equilibria. The other type of refinement is more natural for pure strategy equilibria in games on continuous actions spaces. In many settings—particularly those relevant for machine learning and learning-based systems—the cost functions of players are *non-convex* and hence, we consider *local* notions of equilibria. That is, equilibria such as Nash or Stackelberg defined in a local neighborhood. This will be covered in more detail in Chapter 4. As with correlated equilibria, the local notions of equilibria are important because again many learning algorithms only have local guarantees on convergence due to the non-convex optimization landscape in which players are interacting.

For instance, we can define a δ-local Nash equilibrium as follows. Let $B_\delta(x_i)$ be the δ–radius ball centered at x in the Euclidean space $\mathbb{R}^{m_i}$ endowed with the usual metric.

Definition 2.15. A joint action profile x^* is an local Nash equilibrium if there exists a $\delta > 0$ such that

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*),\ \forall x_i \in B_\delta(x_i^*)$$

Approximate notions of Equilibria. Yet, another refinement arises due to approximation of the inequalities in the definition of equilibrium concepts such as Nash or Stackelberg. For instance, we can define an ε-approximate Nash equilibrium as follows. Let $B_\delta(x_i)$ be the δ–radius ball centered at x in the Euclidean space $\mathbb{R}^{m_i}$ endowed with the usual metric.

Definition 2.16. Give $\varepsilon, \delta > 0$, a joint action profile x^* is an (ε, δ)-approximate Nash equilibrium if

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*) + \varepsilon, \ \forall x_i \in B_\delta(x_i^*)$$

Chapter 3

Games on Finite Action Spaces

3.1 Introduction

Normal form games in which the action spaces X_i for each player $i \in \mathcal{N}$ is itself a finite set are called *finite games* or *normal form games on finite action spaces.* In this case, we can define or represent the action space for a player i as an index set as follows:

$$X_i = \{1, \ldots, m_i\}$$

where $m_i = |X_i|$ is the number of actions available to player i.

The utility of player i is still a mapping from the joint action space X to a real number. For this class of games there are two relevant types of equilibrium concepts: *pure strategy/action* equilibria and *mixed strategy/action* equilibria. Both are defined as in the previous chapter for a Nash equilibrium. That is,

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*),\ \forall x_i \in X_i$$

where if we are considering pure strategy Nash equilibria then $x_i \in X_i = \{1, \ldots, m_i\}$ and if we are considering mixed Nash equilibria, then $x_i \in \Delta(X_i)$—i.e., the probability simplex on $\{1, \ldots, m_i\}$ which is defined by

$$\Delta(X_i) = \left\{ x_i : \sum_{j=1}^{m_i} x_{ij} = 1,\ 0 \leq x_{ij} \leq 1 \right\}$$

The use of x_i for both pure and mixed strategies is of course an abuse of notation in some sense, however, for $x_i \in \Delta(X_i)$, the strategy

$$x_i = (0, \ldots, 0, \underbrace{1}_{j\text{-th entry}}, 0, \ldots, 0)$$

represents a pure strategy where player i puts all the mass on the j-th entry. Since this is an abuse of notation of sorts, using x_i for both pure and mixed strategies, so we will be very clear what we are referring depending on the context moving forward. When we are discussing mixed strategies, we will use x_i for the mixed strategy and either an index such as $j \in X_i$ or $s_i \in X_i$ for the pure strategy. In the case of mixed Nash equilibria, we can write the utility as

$$u_i(x_i, x_{-i}) = \sum_{j=1}^{m_i} x_{ij} u_i(j, x_{-i})$$

where $u_i(j, x_{-i})$ is the utility player i gets from playing the pure strategy j given that the remaining players are playing the mixed strategy x_{-i}.

One reasonable interpretation for mixed policies is the following. Suppose players are playing a game repeatedly. In each game they choose their actions randomly according to pre-selected mixed policies (independently from each other and independently from game to game) and their goal is to minimize/maximize the cost/reward averaged over all the games played.

Moreover, by introducing mixed policies, we essentially enlarged the action spaces for both players and a consequence of this enlargement is that saddle-point equilibria now always exist. Indeed, this is exactly Nash's celebrated theorem provides existence in the class of finite games.

Theorem 3.1 (Nash '51)**.** Every game with a finite number of players and action profiles has at least one Nash equilibrium.

The proof of this result takes time to go through, and we do not discuss it here except to mention that it is typically achieved by appealing to a fixed-point theorem from mathematics, such as those due to Kakutani and Brouwer. Nash's theorem depends critically on the availability of mixed strategies to the agents. However, this does not mean that an agent *plays* a mixed-strategy Nash equilibrium: *Do players really sample probability distributions in their heads?* Some people have argued that they really do, and there is some empirical backing of this.

3.2 Matrix Games: General Sum

A finite normal form game in which there are two players can be described conveniently in a table or matrix form. When there are multiple players (here meaning more than two), the finite normal form game can be represented as a tensor. If it is further the case that there are more than two players and the game is such that it can be written as pairwise games between players, then we refer to such games as *polymatrix games* or *network matrix games.*

For simplicity we will focus on two-player bi-matrix games. Each player has a a finite set of actions:

- player 1 has m actions, indexed by $X_1 = \{1, \ldots, m\}$
- player 2 has n actions, indexed by $X_2 = \{1, \ldots, n\}$

The outcome of the game for player 1 is quantified by an $m \times n$ matrix $A = [a_{ij}]$, and similarly, for player 2, the outcome is quantified by an $n \times m$ matrix $B = [b_{ji}]$. In particular, the entry a_{ij} of the matrix provides the outcome of the game when player 1 selects an action $i \in \{1, \ldots, m\}$, and player 2 selects an action $j \in \{1, \ldots, n\}$. The utilities[1] of the players are then written as

$$u_1(x_1, x_2) = x_1^\top A x_2 \quad \text{and} \quad u_2(x_1, x_2) = x_1^\top B^\top x_2.$$

One can imagine that player 1 selects a row of A and player 2 selects a column of $B^\top$.

The tuple $\mathcal{G} = (A, B)$ denotes the bimatrix game in which player 1 has utility $u_1(x_1, x_2) = x_1^\top A x_2$ and player 2 has utility $u_2(x_1, x_2) = x_1^\top B^\top x_2$.

[1]Note that different references treat the payoff of the "column" player (i.e. player 2) differently in that they may define B using the transpose of the above. So take care when reading other references to understand their notion.

Definition 3.2. A joint action profile $(x_1^*, x_2^*) \in \Delta(X_1) \times \Delta(X_2)$ is a Nash equilibrium for the game $\mathcal{G} = (A, B)$ if

$$(x_1^*)^\top A x_2^* \geq x_1^\top A x_2^*, \ \forall\, x_1 \in \Delta(X_1)$$

and

$$(x_1^*)^\top B^\top x_2^* \geq (x_1^*)^\top B^\top x_2, \ \forall\, x_2 \in \Delta(X_2)$$

Note that if the matrices A and B represent costs to players, then the inequalities in the above definition simply flip.

Example 3.3 (General Sum Bi-matrix Game)**.** Consider the following general sum game:

P1\P2	L	R
U	$(5,1)$	$(0,0)$
D	$(4,4)$	$(1,5)$

The matrix form of this game is defined by the following matrices:

$$A = \begin{bmatrix} 5 & 0 \\ 4 & 1 \end{bmatrix} \quad \text{and} \quad B^\top = \begin{bmatrix} 1 & 0 \\ 4 & 5 \end{bmatrix}$$

One can easily see that (U, L) and (D, R) are the two pure strategy NE of the game.

For a purely mixed strategy, recall that if one player mixes, then the other player should be indifferent between the outcomes of its actions since otherwise it would put all the mass on the preferred action. We can express this in terms of the matrices. Let $a_i^\top$ be the i-th row of A and b_j be the j-th column of $B^\top$. Then, if player 1 uses the mixed strategy $x_1 = (p, 1-p)$, it should be the case for player 2 that it is indifferent between actions L and R. That is,

$$x_1^\top b_1 = x_1^\top b_2 \iff p + 4(1-p) = 0p + 5(1-p) \iff p = \tfrac{1}{2}$$

Similarly, if player 2 uses the mixed strategy $x_2 = (q, 1-q)$, it should be the case that player 1 is indifferent between U and D. That is,

$$a_1^\top x_2 = a_2^\top x_2 \iff 5q + 0(1-q) = 4q + (1-q) \iff q = \tfrac{1}{2}$$

Thus, the unique Nash is $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and the payoff to players is $(5/2, 5/2)$.

This example provides some intuition for how we will ultimately pose finding Nash in bimatrix games as an optimization problem.

3.3 Bi-Matrix Games: Zero Sum

It is even simpler in the zero-sum setting since only a single matrix is needed to define the game. In particular, player 1 wants to minimize the outcome $A(i,j)$ and player 2 wants to maximize $A(i,j)$. For simplicity, since there are two players, in this section we will use $x \in \mathcal{X} = \Delta(X)$ with $X = \{1, \ldots, m\}$ for player 1's mixed strategy and $y \in \mathcal{Y} = \Delta(Y)$ with $Y = \{1, \ldots, n\}$ for player 2's mixed strategy. Then $A \in \mathbb{R}^{m \times n}$ and the game is defined by payoff matrices A and $B = -A^\top$ where player 1 has cost

$$f(x, y) = x^\top A y$$

and player 2 has cost

$$-f(x, y) = -x^\top A y$$

Example 3.4 (Recall Matching Pennies)**.** Consider two players: player 1 (Even) and player 2 (Odd). Each player has a penny and must secretly turn the penny to heads or tails. The players then reveal their choices simultaneously where the rules for outcomes are as follows:

- If the pennies match (both heads or both tails), then Even keeps both pennies, so wins one from Odd (+1 for Even, -1 for Odd).
- If the pennies do not match (one heads and one tails) Odd keeps both pennies, so receives one from Even (-1 for Even, +1 for Odd).

We can write this in matrix form as follows. Player 1 has payoff matrix

$$A = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

where we point out that this defines the *cost* to player 1.[2] Player 2 has payoff matrix

$$B^\top = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = -A$$

We saw before that this game has no pure strategy equilibrium and instead has a mixed strategy equilibrium at $(1/2, 1/2)$ for both players. We can find the equilibria by using similar reasoning as we did in Chapter 2: if a player is playing a mixed strategy in response to a mixed strategy, then they should be indifferent to the actions available to them since otherwise they would put all the mass on the action that gives the best outcome and hence, be playing a pure strategy.

Indeed, with $y = (q, 1-q)$ we have

$$a_1^\top y = a_2^\top y \iff -q + (1-q) = q - (1-q) \iff q = \tfrac{1}{2}$$

and similarly, with $x = (p, 1-p)$, we have

$$x^\top b_1 = x^\top b_2 \iff p - (1-p) = -p + (1-p) \iff p = \tfrac{1}{2}$$

Note that here the columns of $B^\top$ are the columns of $-A$.

The definition of a Nash equilibrium in zero-sum bimatrix games simplifies in some sense, and there is a connection to saddle points.

Recall Definition 3.2 that we defined a mixed Nash equilibrium for the game $\mathcal{G} = (A, B)$ where the matrices defined the players payoffs in terms of utility (as opposed to cost). Treating the players as cost minimizers (i.e., letting player 1 have cost $x^\top Ay$ and player 2 have cost $x^\top B^\top y$), then a joint action profile $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is a (mixed) Nash equilibrium for the cost minimization game $\mathcal{G} = (A, B)$ if

$$(x^*)^\top Ay^* \le x^\top Ay^*, \ \forall\ x \in \mathcal{X} \tag{3.1}$$

and

$$(x^*)^\top B^\top y^* \le (x^*)^\top B^\top y, \ \forall\ y \in \mathcal{Y} \tag{3.2}$$

[2] Typically you will see matching pennies and other games written such that A above is taken to be $-A$ (defining a utility for player 1), but since we want to be consistent with having player 1 as the minimizer, we define A to be a cost matrix for player 1.

Since in zero-sum settings we have $B = -A^\top$, the second inequality becomes

$$-(x^*)^\top Ay^* \le -(x^*)^\top Ay \iff (x^*)^\top Ay^* \ge (x^*)^\top Ay, \ \forall\, y \in \mathcal{Y}$$

If we combine this inequality with (3.1), we get the definition of (mixed) Nash for zero sum finite games and can immediately see its connection to saddle points.

Definition 3.5. A joint action profile $(x_1^*, x_2^*) \in \Delta(X_1) \times \Delta(X_2)$ is a Nash equilibrium for the zero-sum (cost minimization) game $\mathcal{G} = (A, -A^\top)$ if

$$(x^*)^\top Ay \le (x^*)^\top Ay^* \le x^\top Ay^*, \ \forall\, x \in \mathcal{X}, \ y \in \mathcal{Y}$$

The above definition shows that (x^*, y^*) is a saddle point of the function $x^\top Ay$ defined over $\mathcal{X} \times \mathcal{Y}$. Indeed, recall that a vector (x^*, y^*) is a saddle point if

$$\max_{y\in\mathcal{Y}} (x^*)^\top Ay = (x^*)^\top Ay^* = \min_{x\in\mathcal{X}} x^\top Ay^* \tag{3.3}$$

The minmax inequality is a useful tool as we will see not only for finite games but also continuous ones.

Proposition 3.6 (Min-max inequality)**.** Given any function $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$, the following inequality holds:

$$\sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} f(x,y) \le \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x,y) \tag{3.4}$$

Proof. For every $y' \in X$,

$$\inf_{x\in\mathcal{X}} f(x,y') \le \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x,y)$$

Now, just take the supremum over $y' \in \mathcal{Y}$ on the left hand side and we done. ∎

Note that for utility (or cost) functions in bimatrix games, we can replace sup and inf with max and min, respectively, when $f(x,y) = x^\top Ay$ is a function on the joint mixed strategy spaces $\mathcal{X} \times \mathcal{Y}$. That is,

$$\max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} x^\top Ay \le \min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} x^\top Ay \tag{3.5}$$

Now reflecting back to (3.3), we see that it implies that

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} x^\top Ay \le \max_{y\in\mathcal{Y}} (x^*)^\top Ay = (x^*)^\top Ay^* = \min_{x\in\mathcal{X}} x^\top Ay^*) \le \max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} x^\top Ay$$

Hence, combining this with (3.5), we see that a mixed joint action profile (x^*, y^*) is a Nash equilibrium if and only if

$$V(A) = (x^*)^\top Ay^* = \min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} x^\top Ay = \max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} x^\top Ay \tag{3.6}$$

Moreover, we refer to

$$(x^*)^\top Ay^*$$

as the *value of the game*. The value of the game for zero-sum games is an important construct.

3.3.1 Security Values in Zero-sum Games

In this subsection, we discuss what are known as security policies and security values in zero-sum games as a warm up to the minmax theorem of von Neumann which informally says that each player receives a payoff that is equal to both his maxmin value and his minmax value in a Nash equilibrium. Security values correspond to minmax and maxmin values of the game. Let's formalize this.

A *secure* or *risk averse* policy makes choices that are guaranteed to produce the *best outcome* against any choice made by the other player (rational or not).

Given a game $\mathcal{G} = (A, -A^\top)$, the *average* (because we are taking the expected utility given the mixed strategies of players) *security level* for player 1 is defined by

$$\bar{V}(A) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top A y \tag{3.7}$$

and the corresponding *mixed security policy* for player 1 is any x^* that achieves $\bar{V}(A)$:

$$\max_{y \in \mathcal{Y}} (x^*)^\top A y = \bar{V}(A) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top A y \tag{3.8}$$

Analogously, the security level for player 2 is

$$\underline{V}(A) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top A y \tag{3.9}$$

and the security policy is

$$\min_{x \in \mathcal{X}} x^\top A y^* = \underline{V}(A) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top A y \tag{3.10}$$

Proposition 3.7. Consider any zero-sum finite game $\mathcal{G} = (A, -A^\top)$. The following hold:

a. The average security levels are well defined and unique.

b. Both players have mixed security policies (not necessarily unique).

c. The average security levels always satisfy

$$\underline{V}(A) \leq \bar{V}(A) \tag{3.11}$$

The first property follows from an application of Weierstrass' Theorem.[3] Indeed, given that we are minimizing/maximizing a continuous function over a compact set (i.e., bounded and closed) and Weierstrass' Theorem guarantees that such a minimum/maximum always exists at some point in the set.

Then, given that the min and max in the definition of $\underline{V}(A)$ are achieved at specific points in $\mathcal{X}$ and $\mathcal{Y}$, respectively, we get directly that y^* can be used in a security policy for player 2. We can

[3] From calculus, one may recall the extreme value theorem which states the following: a real-valued function $f \in C([a, b], \mathbb{R})$ must attain a maximum and a minimum, each at least once. This can be extended to metric spaces and general topological spaces for that matter.

Theorem 3.8. Suppose that X is a compact set and $f \in C(X, \mathbb{R})$. Then f is bounded and there exists $x, x' \in X$ such that $f(x) = \sup_{z \in X} f(z)$ and $f(x') = \inf_{z \in X} f(z)$

reason in a similar way for player 1. Hence, this shows that the second point in Proposition 3.7 holds.

The inequality in the last property follows from the minmax inequality discussed above. Moreover, it is not hard to show that

$$\bar{V}(A) \leq \min_{i \in X} \max_{j \in Y} A(i,j)$$

and

$$\max_{j \in Y} \min_{i \in X} A(i,j) \leq \underline{V}(A)$$

That is, the mixed security levels are bounded by the *pure security levels*—i.e., equivalent security level concept considering only pure strategies $(i,j) \in X \times Y$.

Now, reflecting back to (3.6), we can state the existence of a mixed Nash equilibrium in terms of the security levels of the game.

Theorem 3.9. Given a zero-sum finite matrix game $\mathcal{G} = (A, -A^\top)$, a mixed Nash equilibrium exists if and only if

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top Ay = \underline{V}(A) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top Ay = \bar{V}(A)$$

This is actually known as the Minmax Theorem. We will see its proof in the next section.

Before getting into the next example, we need the following lemma about optimizing linear functions over the simplex.

Proposition 3.10. Consider the simplex $\Delta(X) = \{x \in \mathbb{R}^m : \sum_i x_i = 1, \; x_i \geq 0, \forall i\}$. Suppose we aim to optimize $f(x) = \sum_{i=1}^m a_i x_i$. Then,

$$\min_{x \in \Delta(X)} f(x) = \min_{i \in \{1,\ldots,m\}} a_i, \quad \max_{x \in \Delta(X)} f(x) = \max_{i \in \{1,\ldots,m\}}$$

Example 3.11 (Rock-Paper-Scissors (Ro-Sham-Bo))**.** Another classical game in game theory is rock-paper-scissors. For the uninitiated, in its simplest form, two players shake their fists and "throw" (with their hand) a symbol for rock (fist), scissors (two fingers out), or paper (flat hand). The outcome of the game are determined by the rules which state that paper beats rock (it can cover it), rock beats scissors (it can smash it), and scissors beats paper (it can cut it). The pay off matrix is given by

$$A = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

Suppose that player 1 uses mixed policy $x = (x_1, x_2, x_3)$ and player 2 uses mixed policy $y = (y_1, y_2, y_3)$. Then,

$$\underline{V}(A) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top Ay = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x_1(y_2 - y_3) + x_2(y_3 - y - 1) + x_3(y_1 - y_2) \tag{3.12}$$

$$= \max_{y \in \mathcal{Y}} \min\{(y_2 - y_3), (y_3 - y_1), (y_1 - y_2)\} \tag{3.13}$$

where the last equality comes from Proposition 3.10, and this is clearly maximized at $y_1 = y_2 = y_3 = \frac{1}{3}$. This means that $\underline{V}(A) = 0$ which gives the mixed security policy of $y^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for player 2. One can similarly argue that $\bar{V}(A) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top Ay = 0$ and $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

Perhaps unsurprisingly, $V(A) = \underline{V}(A) = \bar{V}(A) = 0$ and thus we have a mixed strategy equilirbium:

$$(x^*, y^*) = ((\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}), (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}))$$

3.3.2 Minmax Theorem

In this section, we will see why the Minimax Theorem, which we restate below, holds.

Theorem 3.12 (Minimax Theorem of von Neumann 1928)**.** For any finite, two-player zerosum game, in any Nash equilibrium each player receives a payoff that is equal to both his maxmin value and his minmax value. That is,

$$\underline{V}(A) = \max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} x^\top Ay = \min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} x^\top Ay = \bar{V}(A)$$

This is a highly important theorem in that it demonstrates that maxmin strategies, minmax strategies and Nash equilibria coincide in two-player, zero-sum games.[4]

To prove the theorem we need a bit of background material. A hyperplane in $\mathbb{R}^n$ is a set of the form

$$\{x : \ a^\top x = b\}$$

where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$—i.e., the set of points with a constant inner product to a given vector a, or as a hyperplane with normal vector a. The constant $b \in \mathbb{R}$ determines the offset of the hyperplane from the origin. From this perspective we can rewrite the hyperplane in the form

$$P = \{x \in \mathbb{R}^n : \ a^\top(x - x_0) = 0\},$$

where $x_0 \in \mathbb{R}^n$ is a point that belongs to the hyperplane—i.e., any point satisfying $a^\top x_0 = b$—and $a \in \mathbb{R}^n$ is the normal (vector) to the hyperplane.

A hyperplane divides $\mathbb{R}^n$ into two *halfspaces*. A closed halfspace is a set of the form

$$H = \{x : \ a^\top x \leq b\} = \{x : \ a^\top(x - x_0) \leq 0\}$$

where $a \neq 0$. The open halfspace replaces $\leq$ with $<$.

Theorem 3.13 (Separating Hyperplane Theorem[5])**.** Consider $K, C \subset \mathbb{R}^n$ to be non-empty convex set such that $C \cap K = \emptyset$. There exists $a \neq 0$ and $x_0 \in \mathbb{R}^n$ such that $a^\top(x - x_0) \leq 0$ for all $x \in C$ and $a^\top(x - x_0) \geq 0$ for all $x \in K$. The hyperplane $\{x : \ a^\top(x - x_0) = 0\}$ is called the separating hyperplane for C and K.

Another way to state this theorem is as follows: for every convex set $C \subset \mathbb{R}^n$ and point $x_0 \notin C$, there exists a hyperplane $P = \{x : \ a^\top(x - x_0) = 0\}$ that contains x_0 but does not intersect C.

Lemma 3.14 (Theorem of the Alternative[6])**.** Given $A \in \mathbb{R}^{m\times n}$ and a vector $b \in \mathbb{R}^m$, either there exists a vector x such that $Ax \leq b$ and $x \geq 0$ or there exists a vector y such that $A^\top y \geq 0$, $y \geq 0$ and $b^\top y < 0$, but not both.

[4] This is not true in games on continuous action spaces unless they are highly structured, for example convex games on convex, compact strategy spaces.

[5] cf. [10], Section 2.5.1, pgs. 46-49

[6] cf. Section 5.8.3 [10]

To understand what this lemma is saying, consider the primal and dual optimization problems given by

$$\max_x \{0^\top x \mid x \geq 0,\ Ax \leq b\} \tag{P}$$

and

$$\min_y \{b^\top y \mid y \geq 0,\ A^\top y \geq 0\} \tag{D}$$

respectively.[7] What this lemma is essentially trying to say is that either the primal problem is feasible, or the dual has a feasible solution with negative objective function but not both. Theorems of alternatives[8] essentially look at the problem of determining feasibility of a system of inequalities and equalities and the way this is done is using Lagrangian duality. Here, we have a simple set of linear inequalities $x \geq 0$ and $Ax \leq b$, and we can use duality to show that either there is an x that is feasible for this set of inequalities (i.e., satisfies them) or there has to be a feasible point for the dual problem with negative objective. Let us now see why this lemma holds.

Proof Sketch. Examining (P) we see that since the objective function is zero for all x, the primal cannot be unbounded. Examining (D) we see that the dual is always feasible since we can take $y = 0$.

This means that the primal can only be infeasible or have an optimal solution,[9] while the dual can only be unbounded or have an optimal solution.

By strong duality[10], if either one has an optimal solution, then both do. Hence, the only cases are the following:

1. Both primal and dual have optimal solutions. In particular the primal is feasible, which is the first alternative.
2. The primal is infeasible and the dual is unbounded. In this case the dual has a family of feasible solutions with objective function tending to $-\infty$ which means that the dual must have a feasible solution with $b^\top y < 0$. This is the second alternative.

Now we have to see why these two are mutually exclusive. This follows from weak duality: suppose x is feasible for the primal and y feasible for the dual so that by weak duality $0^\top x \leq b^\top y$ but this contradicts $b^\top y < 0$. ■

Using these classical results, we now proceed to the prove the Minimax Theorem.

Proof Sketch of the Minmax Theorem. Recall that we saw earlier that

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top A y = \underline{V}(A) \leq \overline{V}(A) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top A y$$

If the inequality were strict (meaning the Minmax Theorem does not hold), then there would be a constant c such that

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top A y = \underline{V}(A) < c < \overline{V}(A) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top A y$$

---

[7] For a reminder on the basics of primal-dual forms of linear programs, please see [10]. In particular, for the linear program $\max_x \{c^\top x \mid x \geq 0,\ Ax \leq b\}$ its dual is easily shown to be $\min_y \{b^\top y \mid y \geq c,\ A^\top y \geq 0\}$ and here, we just take c to be the zero vector.

[8] Note that there are several and you may look at Section 5.8 of [10] for more detail on these theorems.

[9] The options here for a linear program are that it can either be unbounded, infeasible, or have an optimal solution.

[10] If the primal has an optimal solution then the dual has an optimal solution too, and the two optima are equal.

So, we need to argue this is not possible. We do this by showing that for any $c \in \mathbb{R}$, we have either

$$c \le \underline{V}(A) \text{ or } \bar{V}(A) \le c$$

From here you may be able to see why we need the theorem of alternatives. Pick arbitrary $c \in \mathbb{R}$. To show that either

$$c \le \underline{V}(A) \text{ or } \bar{V}(A) \le c$$

we use Lemma 3.14 (Theorem of Alternatives) to break the above into the following two cases:

1. There exists $y^* \in \mathcal{Y}$ such that

$$x^\top Ay^* \ge c,\ \forall x \in \mathcal{X} \implies \min_{x\in\mathcal{X}} x^\top Ay^* \ge c \implies \underline{V}(A) = \max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} x^\top Ay \ge c$$

2. There exists $x^* \in \mathcal{X}$ such that

$$(x^*)^\top Ay \le c,\ \forall y \in \mathcal{Y} \implies \max_{y\in\mathcal{Y}} (x^*)^\top Ay \le c \implies \bar{V}(A) = \min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} x^\top Ay \le c$$

Lemma 3.14 is what gives us the existence of y^* and x^* in the above two conditions. One can see this by applying the lemma to the matrix $M = A - c\mathbf{1}$ where $\mathbf{1} \in \mathbb{R}^{m\times n}$ is the matrix with all ones. Then, when the primal feasibility holds in the lemma (first alternative), we know there exists $x^* \in \mathcal{X}$ such that

$$(x^*)^\top (A - c\mathbf{1})y = (x^*)^\top Ay - c \le 0,\ \forall y \in \mathcal{Y}$$

which shows condition 2 above holds. Alternatively, when the dual feasibility with negative objective value (second alternative) holds in the lemma, there exists $y^* \in \mathcal{Y}$ such that

$$x^\top (A - c\mathbf{1})y^* = x^\top Ay^* - c \ge 0,\ \forall x \in \mathcal{X}$$

so that condition 1 above holds.

Now, as noted above if the Minmax Theorem did not hold, then we could pick c such that

$$\underline{V}(A) < c < \overline{V}(A)$$

but this would clearly contradict the two conditions 1 and 2 above that we just proved.

■

3.4 Computing Nash Equilibria in Bi-Matrix Games

In this section, we give an overview on how to compute Nash equilibria in bi-matrix games. We take an optimization perspective since it aligns with the optimization perspective we will take when we discuss continuous games. We start with zero-sum games because they are the simplest to work with from this perspective. Indeed we will see that computing Nash can be done by solving a pair of linear programs. After this, we generalize as much as possible to bi-matrix games and then comment on other ways of computing Nash.

Brief Review of Primal-Dual Representations of Linear Programs. Essentially the main tool we use from convex optimization for matrix games is linear programming. Consider the linear program

$$\min\{b^\top x : \ A^\top x \geq c, \ x \geq 0\}$$

Its dual is given by

$$\max\{c^\top y : \ Ay \leq b, \ y \geq 0\}$$

Theorem 3.15 (Weak duality)**.** Consider the primal dual pair above. Then,

1. (Weak duality) $c^\top y \leq b^\top x$ for all feasible x and y.
2. (Strong duality) If the feasible sets are non-empty, then $c^\top y = b^\top x$ for some feasible points.

3.4.1 Zero-Sum Bimatrix Games

The problem of finding a mixed Nash equilibrium of a finite zero-sum game can be written as a pair of linear optimization problems.

To see this, we can use Proposition 3.10 from the proceeding section. Consider a fixed value of $x \in \mathcal{X}$, so that we have

$$\max_{y \in \mathcal{Y}} x^\top A y = \max_{y \in \mathcal{Y}} \sum_{ij} x_i a_{ij} y_j = \max_{y \in \mathcal{Y}} \sum_j \left(y_j \sum_i x_i a_{ij} \right) = \max_j \sum_i x_i a_{ij} = \max_j \{[x^\top A]_j\}$$

Now we can convert the max over a set of numbers on the right hand side to a minimization:

$$\max_j z_j = \min\{v \in \mathbb{R} : \ v \geq z_j \ \forall j\}$$

Hence,

$$\max_{y \in \mathcal{Y}} x^\top A y = \min_v \{v : \ v \geq [x^\top A]_j, \ \forall j\}$$

Then, the minmax value of the game is

$$\overline{V}(A) = \min_{x \in \mathcal{X}} \min_v \{v : \ v\mathbf{1} \geq A^\top x\} \tag{3.14}$$

$$= \min_{v,x} \{v : \ x \geq 0, \ \mathbf{1}^\top x = 1, \ A^\top x \leq v\mathbf{1}\} \tag{3.15}$$

This is an optimization over $m+1$ parameters $(v, x_1, \ldots, x_m)$. Solving this problem we not only obtain the value of the game v^* but also a mixed security policy x^* for player 1.

We can argue similarly for the maximizing player (player 2). Fix a value of $y \in \mathcal{Y}$, so that we have

$$\min_{x \in \mathcal{X}} x^\top A y = \min_{x \in \mathcal{X}} \sum_{ij} x_i a_{ij} y_j = \min_{x \in \mathcal{Y}} \sum_i \left(x_i \sum_j a_{ij} y_j \right) = \min_i \sum_j a_{ij} y_j = \min_i \{[Ay]_i\}$$

Now we can convert the min over a set of numbers on the right hand side to a maximization:

$$\min_i z_i = \max\{v \in \mathbb{R} : \ v \leq z_i \ \forall i\}$$

Hence,

$$\min_{x\in\mathcal{X}} x^\top Ay = \max_v\{v:\ v \leq [Ay]_i,\ \forall i\}$$

so that the maxmin value of the game is

$$\begin{aligned}\underline{V}(A) &= \max_{y\in\mathcal{Y}}\max_v\{v:\ v\mathbf{1} \leq Ay\} &(3.16)\\ &= \max_{v,y}\{v:\ y \geq 0,\ \mathbf{1}^\top y = 1,\ Ay \geq v\mathbf{1}\} &(3.17)\end{aligned}$$

Solving this problem we get the value of the game v^* and a mixed security policy for player 2.

These two problems we wrote down are in fact dual of one another, and by strong duality, we have a solution $x^* = y^*$. This is perhaps a surprising fact, that zero-sum games always have at least one symmetric Nash equilibrium.

Now, if we want all the mixed security policies for each of the players we can use the value of the game to obtain this set:

$$\{x \in \mathbb{R}^m |\ x \geq 0,\ \mathbf{1}^\top x = 1,\ v^*\mathbf{1} \geq A^\top x\}$$

and

$$\{y \in \mathbb{R}^n |\ y \geq 0,\ \mathbf{1}^\top y = 1,\ v^*\mathbf{1} \leq Ay\}$$

Linear programs can be solved in polynomial time (poly in m and n).

Seeing Minmax Theorem from Dual of the Dual. Let us now in a sense repeat what we have seen above in a slightly different way that suggests that the maxmin problem is dual to the minmax problem and hence, for zero-sum games, strong duality gives the celebrated minmax theorem.

Mixed strategies x and y of the two players are nonnegative vectors whose components sum up to one. These conditions can be viewed as linear constraints:

$$\mathcal{X} = \{\mathbf{1}^\top x = 1,\ x \geq 0\} \text{ and } \mathcal{Y} = \{\mathbf{1}^\top y = 1,\ y \geq 0\}$$

Given a fixed $y \in \mathcal{Y}$, a best response of player 1 to y is a vector $x \in \mathcal{X}$ that minimizes the expression $x^\top(Ay)$. That is, x is a solution to the linear program

$$\min\{x^\top(Ay):\ \mathbf{1}^\top x = 1,\ x \geq 0\} \tag{3.18}$$

The dual of this linear program is

$$\max\{v:\ v\mathbf{1} \leq Ay\}$$

Consider a zero-sum game $\mathcal{G} = (A, -A^\top)$. This minimum payoff to player 1 is the optimal value of the linear program in (3.18), which is equal to the optimal value of v in its dual. Player 2 is interested in maximizing v by their choice of y. The constraints of the dual are linear in v and y even if $y \in \mathcal{Y} = \{y:\ \mathbf{1}^\top y = 1, y \geq 0\}$ is treated as the variable. Hence, a maxmin strategy y for player 2—i.e., *maximizing the minimum* amount they receive—is a solution to the linear program given by

$$\max_{v,y}\{v:\ \mathbf{1}^\top y = 1,\ Ay - v\mathbf{1} \geq 0,\ y \geq 0\} \tag{3.19}$$

It is not hard to see that this linear program gives us a maxmin strategy for the y player.

Now, the dual of this linear program in (3.19) has variables x and u corresponding to the primal constraints $\mathbf{1}^\top y = 1$ and $Ay - v\mathbf{1} \geq 0$, respectively, and is given by

$$\min_{u,x}\{u:\ \mathbf{1}^\top x = 1,\ u\mathbf{1} - A^\top x \geq 0,\ x \geq 0\} \tag{3.20}$$

This linear program gives us a minmax strategy for the x player, since they are *minimizing the maximum* amount they will pay.

Theorem 3.16 (von Neumann Min-Max via Duality [44])**.** A zero-sum cost minimization game $\mathcal{G} = (A, -A^\top)$ has the equilibrium (x^*, y^*) if and only if (u^*, y^*) is an optimal solution to the linear program (3.19) and (v^*, x^*) is an optimal solution to its dual (3.20). Moverover, u^* is the minmax value of the game and v^* is the maxmin value of the game such that $u^* = v^*$.

It is also a very interesting fact that linear programs can be expressed as zero-sum games (cf. [1, 13]).

3.4.2 Computing Nash via Optimization in General Sum Bi-Matrix Games

Similar to the zero-sum case, finding Nash in bi-matrix general sum games can be approached via tools from optimization. In the next chapter, we will make a similar connection between optimization and computation of Nash for continuous games. Such a formulation of equilibrium conditions as an optimization or feasibility problem can be used to infer agents' utility functions and can be used in constructing an an inverse formulation for designing feedback mechanisms to induce a Nash equilibrium. It also leads to several natural learning dynamics which we will see in Chapter 5. So this optimization framework is quite important.

Our focus will remain on two player matrix games since this allows us to introduce concepts more easily, however, it should be noted that many things we discuss can be extended in some manner to network or polymatrix games, albeit with worse computational complexity.

Recall that the joint strategy space is the cross product of of the following two simplexes:

$$\left\{x:\ x = (x_1, \ldots, x_n),\ \sum_i x_i = 1\right\} \times \left\{y:\ y = (y_1, \ldots, y_m),\ \sum_i y_i = 1\right\} = \mathcal{X} \times \mathcal{Y}$$

Let us start by reasoning about points on the interior of $\mathcal{X} \times \mathcal{Y}$ (inner or totally mixed strategies).

Consider a cost minimization bi-matrix game $\mathcal{G} = (A, B)$. Let a_i denote the *rows* of A and let b_j denote the *columns* of $B^\top$. Given a mixed strategy $x \in \mathcal{X}$, we say that $j \in X = \{1, \ldots, m\}$ is in the support of x if $x_j > 0$. We use the notation $\mathrm{supp}(x)$ to denote the support—i.e.,

$$\mathrm{supp}(x) = \{j \in X:\ x_j > 0\}$$

Recall the best response notation: $\mathrm{BR}_i(x_{-i})$ denotes player i's best response map and it returns the set of strategies which are a best response to x_{-i}. If a joint mixed strategy (x^*, y^*) is a Nash equilibrium, then for any $j \in X$

$$j \in \mathrm{BR}_1(y^*) \iff x_j^* > 0$$

since all pure strategies in the support of a Nash equilibrium strategy yields the same payoff or otherwise the player would have a strict preference for the pure strategy with the highest payoff. The gives us the following proposition.

Proposition 3.17. A point (x^*, y^*) is a Nash equilibrium of $\mathcal{G} = (A, B)$ if and only if every pure strategy in the support of x^* is a best response to y^*. That is, (x^*, y^*) is a Nash equilibrium if and only if for all pure strategies in $(i, j) \in X \times Y$,

$$x_i > 0 \implies a_i^\top y = \min_{\ell \in X} a_\ell^\top y \tag{3.21}$$

$$y_j > 0 \implies x^\top b_j = \min_{\ell \in Y} x^\top b_\ell \tag{3.22}$$

The above proposition essentially captures the fact that all pure strategies in the support of a Nash equilibrium yield the same payoff which is also greater than or equal to the payoffs for strategies outside the support. That is, for all $j, k \in \text{supp}(x^*)$ we have

$$f_1(j, y^*) = f_1(k, y^*),$$

where $f_1(x, y)$ is player 1's cost function and there is a slight abuse of notation in using j and k (pure strategies) to denote the mixed strategies that put all the probability mass on the j-th (respectively, k-th) entry of x. In addition, for all $j \in \text{supp}(x^*)$ and $\ell \notin \text{supp}(x^*)$, we have

$$f_1(j, y^*) \leq f_1(\ell, y^*).$$

Similar conclusions can be drawn about player 2's mixed Nash equilibrium strategy y^*. In other words, for purely mixed strategies (where the support set contains every pure strategy) we have that

$$(x^*, y^*) \text{ is a Nash equilibrium} \iff \begin{cases} a_1^\top y^* & = a_i^\top y^* \quad i = 2, \ldots, m \\ (x^*)^\top b_1 & = (x^*)^\top b_j \quad j = 2, \ldots, n \\ \sum_{i=1}^n x_i & = 1 \\ \sum_{j=1}^m y_j & = 1 \end{cases}$$

This is a set of linear equations and can be solved efficiently.[11]

Not all games have purely mixed Nash strategies. Let us now formulation a naïve or brute force extension of the purely mixed strategy computation strategy for cases where we do not have purely mixed strategies. A mixed strategy profile $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is a Nash equilibrium with $\text{supp}(x) \subseteq X$ and $\text{supp}(y) \subseteq Y$ if and only if

$$\begin{cases} u = a_i y & \forall\, i \in \text{supp}(x) \\ u \geq a_i y & \forall\, i \notin \text{supp}(x) \\ v = x^\top b_j & \forall\, j \in \text{supp}(y) \\ v \geq x^\top b_j & \forall\, j \notin \text{supp}(y) \\ x_i = 0 & \forall\, i \notin \text{supp}(x) \\ y_j = 0 & \forall\, j \notin \text{supp}(y) \end{cases}$$

The issue with this computation method is that is requires finding the *right* support sets $\text{supp}(x)$ and $\text{supp}(y)$. There are 2^{m+n} different support sets; hence, this processes is exponential in computation time.

[11] For games with more than 2 players, a similar line of reasoning leads to a set of polynomial equations.

An Optimization Approach

Instead of taking the approach of constructing the support sets (which works well for games with purely mixed strategies, if we know a priori that such an equilibrium exists), we take an optimization approach. Indeed, this approach to finding equilibrium strategies for a bi-matrix game essentially transforms the problem into a nonlinear (in fact, a bilinear) programming problem.

Consider a general sum game $\mathcal{G} = (A, B)$. Observe that a best response of player 1 against a mixed strategy y of player 2 is a solution to the linear program (3.18), which we recall here for ease of access:

$$\min\{x^\top (Ay) : \ \mathbf{1}^\top x = 1, \ x \geq 0\}$$

The dual of this linear program is

$$\max\{v : \ Ay \geq v\mathbf{1}\}$$

By strong duality, a feasible solution x is optimal if and only if there is a dual solution v fulfilling $Ay \geq v\mathbf{1}$ and $x^\top (Ay) = v$—that is, $x^\top (Ay) = x^\top \mathbf{1} v$ which itself is equivalent to

$$x^\top (\mathbf{1}v - Ay) = 0$$

Since the vectors x and $\mathbf{1}v - Ay$ are both non-negative, the above equality says that they have to be what is called *complementary* in that both cannot have positive components in the same position. This characterization of an optimal primal-dual pair of feasible solutions is known as *complementary slackness* in linear programming.

Since x has at least one positive component, the respective component of $\mathbf{1}v - Ay$ is zero, and v is the maximum of the components of Ay. Indeed, any pure strategy $i \in X$ is a best response to y if and only if the i-th component of the slack vector $\mathbf{1}v - Ay$ is zero (since otherwise there would be no mass on the strategy i in x_i). This is saying that

$$[\mathbf{1}v - Ay]_i = 0 \iff [x_i > 0 \implies a_i^\top y = \max_{\ell \in X} a_\ell^\top y]$$

Now, for player 2, y is a best response to x if and only if it minimizes $(x^\top B^\top)y$ subject to $y \in \mathcal{Y}$. The dual of this linear program is

$$\max\{u : \ Bx \geq u\mathbf{1}\}$$

A primal-dual pair (u, y) of feasible solutions is optimal if and only if

$$y^\top (\mathbf{1}u - Bx) = 0$$

which is completely analogous to the complementarity problem discussed above for player 1.

Theorem 3.18. The cost minimization game $\mathcal{G} = (A, B)$ has the Nash equilibrium (x^*, y^*) if and only if for some u, v the following hold:

$$\mathbf{1}^\top y = 1, \ \mathbf{1}^\top x = 1, \ x \geq 0, \ y \geq 0 \tag{3.23}$$

$$Ay - v\mathbf{1} \geq 0 \tag{3.24}$$

$$Bx - u\mathbf{1} \geq 0 \tag{3.25}$$

We can further introduce a cost to the above set of inequalities to get a bilinear optimization problem which has as solutions the Nash equilibria for the game.

Proposition 3.19. A mixed profile $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is a Nash equilibrium of the bimatrix cost minimization game $\mathcal{G} = (A, B)$ if and only if there exists (p, q) such that (x, y, u, v) is a solution to

$$\begin{aligned} \min \quad & x^\top Ay + x^\top B^\top y - u - v \\ \text{s.t.} \quad & Ay - v\mathbf{1} \geq 0 \\ & Bx - u\mathbf{1} \geq 0 \\ & \mathbf{1}^\top y = 1,\ y \geq 0,\ \mathbf{1}^\top x = 1,\ x \geq 0 \end{aligned} \tag{P-1}$$

Lemma 3.20. Every mixed strategy Nash equilibrium has a zero optimal value to (P-1).

Proof. Recall that if a point (x^*, y^*) is a Nash equilibrium for $\mathcal{G} = (A, B)$ then x^* is a best response to y^* and vise versa. That is,

$$x^* \in \arg\min\{x^\top Ay^* |\ x \in \mathcal{X}\}$$

The dual of this problem is

$$\max_v \{v|\ Ay \geq \mathbf{1}v\}$$

Feasible points are optimal if and only if the two objective points are equal $(x^*)^\top Ay^* = v^*$. A similar argument holds for $(x^*)^T B^\top y^* = u^*$. Thus,

$$(x^*)^\top Ay^* + (x^*)^\top By^* - v^* - u^* = 0$$

■

Applying this lemma, we see that (x, y) are best responses to one another if and only if

$$x^\top (Ay - \mathbf{1}v) = 0,\ y^\top (Bx - \mathbf{1}v) = 0$$

Proof of Proposition 3.19. Suppose that (x^*, y^*) is a Nash equilibrium. For any feasible solution (x, y, u, v) of (P-1), the constraints imply that

$$x^\top Ay + x^\top B^\top y - u - v \geq 0 \tag{3.26}$$

Suppose that $v^* = (x^*)^\top Ay^*$ and $u^* = (x^*)^\top B^\top y^*$. Then, clearly the vector (x^*, y^*, u^*, v^*) has optimal value zero.

If the vector (x^*, y^*, u^*, v^*) is also feasible, it follows from (3.26) that it is an optimal solution of the optimization problem (since all other feasible solutions have a larger objective value).

Since (x^*, y^*) is a Nash equilibrium,

$$(x^*)^\top Ay^* \leq x^\top Ay^*,\ \forall x \in \mathcal{X}$$

Choosing $x = e_i$—i.e., the i-th unit vector with all zeros except in the i-th entry which has value one—we obtain

$$v^* = (x^*)^\top Ay^* \leq [Ay^*]_i,$$

for each i. This shows that (x^*, y^*, v^*, u^*) satisfies the first constraint $Ay - v\mathbf{1} \geq 0$. A similar argument holds for the second constraint.

This proves that (x^*, y^*, v^*, u^*) is an optimal solution of the optimization problem.

Now, to show the other direction, suppose that (x^*, y^*, v^*, u^*) is an optimal solution of the optimization problem. Since all feasible solutions have non-negative optimal value as argued above, and any mixed strategy Nash equilibrium (which always exists by Nash's theorem) has zero optimal value (by Lemma 3.20), it follows that

$$(x^*)^\top Ay^* + (x^*)^\top By^* - v^* - u^* = 0$$

For any $x \geq 0$ with $\mathbf{1}^\top x = 1$ and $y \geq 0$ with $\mathbf{1}^\top y = 1$, the constraints $Ay - v\mathbf{1} \geq 0$ and $Bx - \mathbf{1}u \geq 0$ imply that

$$x^\top Ay^* \geq v^*,\ y^\top Bx^* \geq u^* \tag{3.27}$$

In particular, we have that

$$(x^*)^\top Ay^* \geq v^*,\ (y^*)^\top Bx^* \geq u^*$$

Combining this with the fact that the objective value is zero, we have

$$(x^*)^\top Ay^* = v^*,\ (y^*)^\top Bx^* = u^*$$

Indeed,

$$0 = (x^*)^\top Ay^* + (y^*)^\top Bx^* - v^* - u^* \geq (x^*)^\top Ay^* - v^* \geq 0$$

where we use the fact that $-u^* \geq -(y^*)^\top Bx^*$ to get the first inequality. The fact that $(y^*)^\top Bx^* = u^*$ can be argued analogously.

Putting this together with (3.27), we get

$$x^\top Ay^* \geq (x^*)^\top Ay^*,\ \forall x \in \mathcal{X}$$

and

$$(x^*)^\top B^\top y \geq (x^*)^\top B^\top y^*,\ \forall y \in \mathcal{Y}$$

■

Chapter 4

Games on Continuous Action Spaces

The second class of normal form games that we review are known as *continuous games* because players possess continuous action spaces.

Perhaps the most natural way to introduce continuous games is using finite games. Indeed, we can think of a normal form game on finite action spaces as a continuous game if we take into consideration *mixed strategies*. For instance, consider the bimatrix game (A, B)—i.e., $f_1(x_1, x_2) = x_1^\top A x_2$ and $f_2(x_1, x_2) = x_2^\top B x_1$. Mixed strategies for this game belong to the simplex $\Delta(X_1) \times \Delta(X_2)$—that is, $x_1 \in \Delta(X_1)$ is a probability distribution on X_1 such that each element in the vector $x_1 = (x_{11}, \ldots, x_{1m_1})$ takes on a value in the continuous interval $[0, 1]$. From this point of view, the game $(x_1^\top A x_2, x_2^\top B x_1)$ can be viewed as a game on continuous actions spaces, albeit constrained action spaces. In this way, mixed strategies over finite games can be thought of as pure strategies in a continuous game over the simplex.

More generally, a game is called a continuous game if the strategy sets are topological spaces and the payoff functions are continuous with respect to the product topology. In particular, in continuous games, players choose actions from a continuous action space, constrained or unconstrained as it may be for the particular rules of the game. In particular, we consider games where players choose from a continuum and the utility/cost functions are continuous.

In the remainder of this chapter, we will introduce the Nash equilibrium concept and refinements (local Nash equilibria) as well as the Stackelberg equilibrium concept and refinements thereof. We will review different types of games including zero-sum, potential, convex, monotone and general sum games. Existence results will be presented (largely without proof), and we will discuss how to compute Nash equilibria taking an optimization approach as we did in the last chapter.

4.1 Nash Equilibrium Definition and Best Response

The definition of a (pure) Nash equilibrium does not change from what we saw for finite games. The only difference is that players' actions are chosen from continuous action spaces.

Definition 4.1. A point x^* is a Nash equilibrium for the continuous game $(f_1, \ldots, f_N)$ if for each $i \in \mathcal{I}$

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*) \ \ \forall \ x_i \in X_i$$

Best response are again useful for characterizing Nash as we have seen in Chapter 2. Towards understanding the best response, we define the rational response set as follows. In an N–person non-zero sum game, let the minimum of the cost function $f_i(x)$ of player i with respect to $x_i \in X_i$ be attained for each $x_{-i} \in X_{-i}$ Then, the set $\mathrm{BR}_i(x_{-i}) \subset X_i$ defined by

$$\mathrm{BR}_i(x_{-i}) = \{\xi \in X_i : f_i(x_i, x_{-i}) \geq f_i(\xi, x_{-i}),\ \forall x_i \in X_i\}$$

is called the *best response set* or *rational reaction set* for player i.

Definition 4.2. A point x^* is a Nash equilibrium if $x_i^* \in \mathrm{BR}_i(x_{-i}^*)$ for each $i \in \mathcal{N}$.

Let's consider a couple examples.

Example 4.3 (Simple Quadratic Game)**.** Consider the game $\mathcal{G} = (f_1, f_2)$ where

$$f_1(x, y) = (x - y)^2 \text{ and } f_2(x, y) = (y - x)^2$$

and $x, y \in \mathbb{R}$. We claim that $(x^*, y^*) = (0, 0)$ is a Nash equilibrium. Indeed,

$$f_1(0, 0) = 0 \leq f_1(x, 0) = x^2, \forall x \in \mathbb{R}$$

and

$$f_2(0, 0) = 0 \leq f_2(0, y) = y^2,\ \forall x \in \mathbb{R}$$

The next example is one of two of the most quintessential continuous games, *Cournot* and *Bertrand* Competition.

Example 4.4 (Cournot Game)**.** In these games, players choose the quantity (respectively, price) of a product given competition for demand. Consider a scenario in which there are two firms. They simultaneously choose any quantity $q_1, q_2 \geq 0$. The price in the market is given by $p(q_1, q_2) = A - q_1 - q_2$ for some constant A. The firms payoff functions are given by

$$u_i(q_i, q_{-i}) = p(q_1, q_2)q_i - c_i q_{-i}$$

for $i = 1, 2$. Consider the optimization problem that each player faces:

$$\max_{q_i} \{u_i(q_i, q_{-i}) |\ q_i \geq 0\} \tag{4.1}$$

Note that the utilities for each player are concave in their individual choice variables q_i. Notice also there are constraints on the problem for each player but they are decoupled.[1] So as in optimization, we will formulate the Lagrangian for each player:

$$L_i(q_i, q_{-i}, \mu_i) = u_i(q_i, q_{-i}) - \mu_i^\top q_i,\ \ \mu_i \geq 0 \tag{4.2}$$

To calculate the equilibrium, we first find critical points of the Lagrangian functions for each player holding the other player fixed. That is, we set their individual partial derivatives to zero as follows:

$$\begin{aligned} \frac{\partial L_1}{\partial q_1} &= A - 2q_1 - q_2 - c_1 - \mu_1 = 0 \\ \frac{\partial L_2}{\partial q_2} &= A - q_1 - 2q_2 - c_2 - \mu_2 = 0 \end{aligned} \tag{4.3}$$

[1] When the constraints are coupled, things become a bit more complicated and we need to introduce the generalized Nash equilibrium concept. We will not do this in this course. We instead will deal with uncoupled constraints or very well behaved constrained sets where things can be uncoupled through clever transformations.

Solving these two equations for q_1, q_2 in terms of μ_1, μ_2 and then using the fact that the Lagrangian for each player is linear in μ_1 and μ_2, respectively, we have the following conditions:

$$\begin{aligned} q_1 &= \tfrac{1}{3}(A + c_2 - 2c_1) + \tfrac{1}{3}(\mu_2 - 2\mu_1) \\ q_2 &= \tfrac{1}{3}(A + c_1 - 2c_2) + \tfrac{1}{3}(\mu_1 - 2\mu_2) \end{aligned} \tag{4.4}$$

Now, we want to minimize the Lagrangian L_i with respect to μ_i and it is linear in μ_i (i.e., the term in L_i is $-\mu_i q_i$) and the original constraint is such that $q_i \geq 0$. Hence, $\mu_i^* = 0$ for each $i = 1, 2$. This gives us the best responses for each player:

$$\begin{aligned} q_1^* &= \tfrac{1}{3}(A + c_2 - 2c_1) \\ q_2^* &= \tfrac{1}{3}(A + c_1 - 2c_2) \end{aligned} \tag{4.5}$$

□

The key point of this example, is that we can find the Nash equilibrium of the game by viewing the problem as a set of coupled optimization problems. We took this similar view in how we formulated the optimization based approach for finding mixed Nash equilibria in bilinear matrix games in the proceeding chapter. This motivates writing down necessary and sufficient conditions for Nash equilibria from this optimization perspective.

Both of the above examples are defined by cost functions that are strongly convex for each of the players. As we will see later this gives some assurances on not only existence but also uniqueness. When it is the case that the cost functions are not convex, we need local notions of equilibria. For a metric space (X, d) with metric d, let $B_\delta(x)$ denote the ball of radius $\mathbb{R}$ centered at x.

Definition 4.5 (Local Nash Equilibrium)**.** A point x^* is a local Nash equilibrium for the continuous game $(f_1, \dots, f_N)$ if for each $i \in \mathcal{N}$, there exists a $\delta_i > 0$ such that

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*) \;\; \forall \; x_i \in B_{\delta_i}(x_i^*).$$

Another way local Nash are defined is as follows: given sets $U_i \subset X_i$ for each $i \in \mathcal{N}$, a point x^* is a local Nash equilibrium on the set $U = U_1 \times \cdots \times U_N$, if for each $i \in \mathcal{N}$,

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*), \;\forall x_i \in U_i.$$

4.2 An Optimization Approach to Nash in Unconstrained Games

We may consider using optimality conditions for each player's individual optimization problem in order to define refinements of the Nash equilibrium concept, and to guide how we may compute such refinements.

Suppose that $X_i \subset \mathbb{R}^{m_i}$ for each i and let

$$X = X_1 \times \cdots \times X_n$$

Here, we will only concern ourselves with full information games and reduce the notation for a game to be the tuple

$$\mathcal{G} = (f_1, \dots, f_N)$$

where each $f_i \in C^2(X, \mathbb{R})$. Consider the unconstrained optimization problem for player i:

$$\min_{x_i \in \mathbb{R}^{m_i}} f_i(x_i, x_{-i}) \tag{4.6}$$

Holding $x_{-i} \in \mathbb{R}^{m-m_i}$ fixed, we can view this simply as an optimization problem in the variable x_i.

A little foray back to calculus. We know from calculus (and nonlinear programming) what the necessary and sufficient optimality conditions are for a extremal point. Recall the definition of a global minima of a (sufficiently smooth) function $h(x)$.

Definition 4.6. A point $x^* \in X$ is a global minima of $h(x)$ if

$$h(x^*) \leq h(x),\ \forall\ x \in X$$

Since $h(x)$ might in general be non-convex in x, we also recall the definition of a local minima.

Definition 4.7. A point $x^* \in X$ is a local minima of $h(x)$ if there exists a $\delta > 0$ such that

$$h(x^*) \leq h(x),\ \forall\ x \in B_\delta(x^*)$$

Moreover, we say that x^* is a local minimum on the set U if the above inequality holds for all $x \in U$.

We know from calculus that a point $x^* \in \mathbb{R}^m$ is a local minimum of

$$\min_{x \in \mathbb{R}^m} h(x)$$

if the following *sufficient* conditions hold:

1. **First order conditions**:
$$Dh(x^*) = 0 \tag{4.7}$$
2. **Second order conditions**:
$$D^2h(x^*) > 0 \tag{4.8}$$

On the other hand, if x^* is a local minimum, then *necessarily* $Dh(x^*) = 0$ and $Dh(x^*) \geq 0$.

Importantly, it is a *generic property* of C^2-functions that all local minima satisfy conditions (4.7) and (4.8) above. This is true not only on Euclidean space but also for sufficiently smooth functions on smooth manifolds (without boundary). Formally, this means that set of functions for which the sufficient conditions above are not also necessary is of measure zero in the class of $C^2(\mathbb{R}^m, \mathbb{R})$ functions.

Coming back to games. Now, let us return to the game setting. First, since $f_i(x_i, x_{-i})$ may also be, in general, non-convex in x_i, we focus on the local Nash equilibrium concept. Recall the optimization problem in (4.6) where we consider for a moment that $x_{-i} \in \mathbb{R}^{m-m_i}$ is fixed. Then this optimization problem over $x_i \in \mathbb{R}^m$ is exactly of the form discussed above and hence, we can write down optimality conditions for each player. Indeed, given $x^*_{-i} \in \mathbb{R}^{m-m_i}$, x^*_i is a local minimum if the following sufficient conditions hold:

1. $D_i f_i(x^*_i, x^*_{-i}) = 0$
2. $D^2_i f_i(x^*_i, x^*_{-i}) > 0$

These individual derivatives $D_i f_i(x)$ are the directions in which a player i can make "differential" adjustments to its strategy in order to improve the cost locally as indicated by the individual curvature $D^2_i f_i(x)$. This observation leads to the following definition of a *differential Nash equilibrium*.

Definition 4.8 (Ratliff, Burden, Sastry, 2013 [45][2]). A strategy $x = (x_1, \ldots, x_N)$ is a *differential Nash equilbrium* of the game $\mathcal{G} = (f_1, \ldots, f_N)$ if $D_i f_i(x) = 0$ and $D_i^2 f_i(x) > 0$ for each $i \in \mathcal{N}$.

The definition of a differential Nash equilibrium is inherently local. Furthermore, differential Nash equilibria are clearly strict local Nash equilibria.

Proposition 4.9 (Ratliff, Burden, Sastry, 2013 [45]). A differential Nash equilibrium of the game $\mathcal{G} = (f_1, \ldots, f_N)$ is a strict local Nash equilibrium.

The conditions of the definition are sufficient for checking if a point is a Nash equilibrium. These look just like first and second order conditions for optimality, but they are not necessary.

Example 4.10. Consider the zero sum game $\mathcal{G} = (f, -f)$ defined by $f(x_1, x_2) = x_1^\top x_2$. Then, $(x_1, x_2) = (0, 0)$ is a local Nash equilibrium since

$$f(0,0) = 0 \leq x_1^\top 0 = 0, \ \forall x_1 \in \mathbb{R}^{m_1}$$

and

$$-f(0,0) = 0 \leq -0^\top x_2 = 0, \ \forall x_2 \in \mathbb{R}^{m_2}$$

□

Yet, following the same construction of necessary conditions for local minima, we can write down necessary conditions for local Nash equilibria.

Proposition 4.11 (Ratliff, Burden, Sastry, 2013 [45]). Suppose x^* is a local Nash equilibrium of the game $\mathcal{G} = (f_1, \ldots, f_N)$. Then, $D_i f_i(x^*) = 0$ and $D_i^2 f_i(x^*) \geq 0$.

One very interesting observation is that unlike strict local minima in single player optimization, a point being a strict local Nash (differential Nash) equilibrium is not sufficient to guarantee that it is isolated. Indeed, it is is easy to construct examples with a continuum of differential Nash equilibria.

Example 4.12 (Continuum of Differential Nash Equlibria.). Consider the game $\mathcal{G} = (f_1, f_2)$ defined by the following cost functions:

$$\begin{aligned} f_1(x_1, x_2) &= \tfrac{1}{2}x_1^2 - x_1 x_2 \\ f_2(x_1, x_2) &= \tfrac{1}{2}x_2^2 - x_1 x_2 \end{aligned} \tag{4.9}$$

Every point on the line $x_1 = x_2$ is a differential Nash equilibrium. Indeed,

$$D_1 f_1(x) = x_1 - x_2 = 0, \ D_2 f_2(x_1, x_2) = x_2 - x_1 = 0 \iff x_1 = x_2$$

and

$$D_1^2 f_1(x) = D_2^2 f_2(x) = 1 > 0, \ \forall x$$

To see these points are local Nash, observe that for any $x_1 = x_2 = y$, we have

$$f_1(y, y) = \frac{1}{2}y^2 - y^2 = -\frac{1}{2}y^2 < x_1^2 - x_1 y, \ \forall x_1 \neq y$$

[2]Note this definition can be extended to a well-defined differential form in settings where the strategy spaces of agents are manifolds (without boundary)

and similarly,

$$f_2(y,y) = \frac{1}{2}y^2 - y^2 = -\frac{1}{2}y^2 < x_2^2 - x_2 y,\ \forall x_2 \neq y$$

This continuum of equilibria is not robust however. That is, even small changes to the game will remove the continuum. Indeed, consider the game

$$\begin{aligned} f_1(x_1,x_2) &= \tfrac{1}{2}x_1^2 - \varepsilon x_1 x_2 \\ f_2(x_1,x_2) &= \tfrac{1}{2}x_2^2 - x_1 x_2 \end{aligned} \tag{4.10}$$

defined for some arbitrary $\varepsilon \neq 1$. Then,

$$D_1 f_1(x) = x_1 - \varepsilon x_2 = 0,\ D_2 f_2(x) = -x_1 + x_2 = 0 \iff x_1 = x_2 = 0$$

That is, the only candidate local Nash equilibrium is now at $x^* = (0,0)$. We can check that it is in fact a Nash equilibrium by check the second order sufficient conditions:

$$D_1^2 f_1(x^*) = 1 > 0,\ D_2^2 f_2(x^*) = 1 > 0$$

This shows that all the differential Nash on the line $x_1 = x_2$ disappear except the one at the origin. □

To rule out a continuum of equilibria, we can check that the game Jacobian, which is given by

$$J(x) = Dg(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}$$

and is the derivative of the vector $g(x) = (D_1 f_1(x), D_2 f_2(x)))$, is non-degenerate at critical points x^*—i.e., $\det(J(x^*)) \neq 0$. Analogous to the case of single player optimization, non-degenerate differential Nash equilibria are generic amongst local Nash equilibria. That is, for all but a measure zero set of games (in the space of games defined by $C^2(X,\mathbb{R})$ functions), local Nash equilibria are non-degenerate differential Nash equilibria.

4.3 Types of Games

There are many classes of games, most categorized by the structure of the cost function. In this section, we will discuss two key types of games.

4.3.1 Potential and Zero-Sum (Hamiltonian) Games

Let's start with the Jacobian of the previous section. This will help us define two important types of games.

For simplicity, let's again consider two players. Recall that the game Jacobian—i.e., the Jacobain of $g(x) = (D_1 f_1(x_1,x_2), D_2 f_2(x_1,x_2))$—is given by

$$J(x) = Dg(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix} \tag{4.11}$$

Consider the asymmetric-symmetric decomposition of $J(x)$:

$$J(x) = \underbrace{\frac{1}{2}(J(x) - J^\top(x))}_{Z(x)} + \underbrace{\frac{1}{2}(J(x) + J^\top(x))}_{P(x)} \tag{4.12}$$

where

$$P(x) = \begin{bmatrix} D_1^2 f_1(x) & \frac{1}{2}(D_{12}f_1(x) + D_{21}^\top f_2(x)) \\ \frac{1}{2}(D_{12}^\top f_1(x) + D_{21}f_2(x)) & D_2^2 f_2(x) \end{bmatrix}$$

and

$$Z(x) = \begin{bmatrix} 0 & \frac{1}{2}(D_{12}f_1(x) - D_{21}^\top f_2(x)) \\ \frac{1}{2}(D_{21}f_2(x) - D_{12}^\top f_1(x)) & 0 \end{bmatrix}$$

These two matrices individual can be thought of as begin the game Jacobians for a *potential game* and a special type of zero sum game where there are no non-coupled terms between the players.

Zero-sum game components. To see this is true let us start with a zero sum $(f, -f)$. In this case, the game Jacobian is

$$J(x) = \begin{bmatrix} D_1^2 f(x) & D_{12}f(x) \\ -D_{12}^\top f(x) & -D_2^2 f(x) \end{bmatrix}$$

and if there are no terms in the function f that depends solely on x_i for $i = 1, 2$, then

$$J(x) = \begin{bmatrix} 0 & D_{12}f(x) \\ -D_{12}^\top f(x) & 0 \end{bmatrix}$$

and it can be easily seen that $Z(x)$ is in the form of this Jacobian. In particular, in consideration of $J(x)$ in (4.11), the game (f_1, f_2) is a zero-sum game if and only if $D_{21}f_2(x) = -D_{21}f_1(x)$. That being said, $Z(x)$ corresponds to a special subclass of zero sum games since more generally, zero sum games do not necessarily have zeros on the block diagonal of the Jacobian.

Potential game component. To understand $P(x)$, we need to recall the definition of a potential game for which we will appeal to the seminal reference by Monderer and Shapley [39]. Potential games are a very nice class of games as they are secretly a a single player optimization problem which is revealed under a transformation of coordinates.

Consider a game $\mathcal{G} = (f_1, f_2)$ defined by cost functions $f_i \in C^2(X, \mathbb{R})$.

Definition 4.13. A function $P : X \to \mathbb{R}$ is a w-potential for the game $\mathcal{G}$ if for every $i \in \mathcal{N}$ and for every $x_{-i} \in X_{-i}$,

$$f_i(x_i, x_{-i}) - f_i(x_i', x_{-i}) = w_i(P(x_i, x_{-i}) - P(x_i', x_{-i})),\ \forall x_i, x_i' \in X_i$$

The game $\mathcal{G}$ is called a w-potential game.

Moreover, if $w = \mathbf{1}$, then we say $\mathcal{G}$ is an exact potential (or just potential) game.

Lemma 4.14. Let P be a real-valued mapping on X. Then P is a potential for $\mathcal{G}$ if and only if $P \in C^1(X, \mathbb{R})$ and

$$\frac{\partial f_i}{\partial x_i} = \frac{\partial P}{\partial x_i},\ \forall i \in \mathcal{N}$$

The following is a well known result from dynamical systems theory and physics, sometime referred to as deRham's theorem.

Theorem 4.15. Let $\mathcal{G} = (f_1, f_2)$ be a game defined by $f_i \in C^2(X, \mathbb{R})$. Then $\mathcal{G}$ is a potential game if and only if

$$D_{12}f_1(x) = D_{12}f_2(x)$$

Moreover, if the cost functions satisfy the above equality and z is an arbitrary but fixed strategy profile in X, then a potential for $\mathcal{G}$ is given by

$$P(x) = \int_0^1 \frac{\partial f_1}{\partial x_1}(\theta(t))\theta_1'(t)dt + \int_0^1 \frac{\partial f_2}{\partial x_2}(\theta(t))\theta_2'(t)dt$$

where $\theta : [0, 1] \to X$ is a piecewise continuously differentiable path in X that connects z to x—i.e., $\theta(0) = z$ and $\theta(1) = x$).

Let's revisit the Cournot game.

Example 4.16. Consider a scenario in which there are two firms. They simultaneously choose any quantity $q_1, q_2 \geq 0$. The price in the market is given by $p(q_1, q_2) = A - q_1 - q_2$ for some constant A. The firms payoff functions are given by

$$u_i(q_i, q_{-i}) = p(q_1, q_2)q_i - c_i q_{-i}$$

for $i = 1, 2$. Consider the optimization problem that each player faces:

$$\max_{q_i}\{u_i(q_i, q_{-i})|\ q_i \geq 0\} \tag{4.13}$$

Now,

$$D_i u_i(q_i, q_{-i}) = A - 2q_i - q_{-i}$$

and

$$\frac{\partial^2 u_1}{\partial q_1 \partial q_2} = -1 = \frac{\partial^2 u_2}{\partial q_1 \partial q_2}$$

Theorem 4.17. Suppose $\mathcal{G} = (f_1, f_2)$ is a potential game defined on compact strategy spaces X_1 and X_2. Then there exists a pure strategy Nash equilibrium of $\mathcal{G}$.

Even without compactness of the strategy spaces, if the cost functions f_i are bounded, we still get existence.

Theorem 4.18. Consider a game $\mathcal{G} = (f_1, f_2)$ defined by bounded cost functions $f_i \in C^2(X, \mathbb{R})$. That is, for each $i \in \mathcal{N}$, there exists a $B_i > 0$, $|f_i(x)| \leq B_i$ for all $x \in X$. Then $\mathcal{G}$ possesses an ε-equlibrium point.

Some Additional Interesting Facts. In this set of lecture notes we do not cover routing games or congestion games, a very important class of games as they are reasonable abstractions for network congestion in communication networks, transportation networks, etc. and for many resource allocation problems. Essentially a congestion game is an abstraction for allocating resources (e.g., drivers) to a finite set of facilities (e.g., roads). At each "facility" there is a congestion cost: the more drivers assigned to a road the costlier it is for those drivers. Monderer and Shapley [39] proved, noting that the result is a reduction from the work of Rosenthal in 1973 [46], that all congestion games are potential games. Moreover, they showed that every finite potential game is isomorphic to a congestion game.

Hamiltonian Games. There is a remarkable relationship between zero-sum, potential and Hamiltonain games. A related game concept is a Hamiltonian game. This is a game such that there exists a Hamiltonian function $H(x)$ such that

$$\begin{bmatrix} D_1 f_1(x) \\ D_2 f_2(x) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{bmatrix} D_1 H(x) \\ D_2 H(x) \end{bmatrix}$$

Zero sum games with no non-interaction components can be Hamiltonian. In particular, unconstrained bimatrix games are Hamiltonian if and only if they are zero sum.

Example 4.19. For example consider the bilinear game

$$f(x, y) = x^\top A y$$

Then,

$$\begin{bmatrix} D_1 f(x, y) \\ -D_2 f(x, y) \end{bmatrix} = \begin{bmatrix} Ay \\ -A^\top x \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & D_{12} f(x, y) \\ -D_{12}^\top f(x, y) & 0 \end{bmatrix} = D \left(\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{bmatrix} D_1 f(x, y) \\ D_2 f(x, y) \end{bmatrix} \right)$$

Hence, $H(x, y) = x^\top A y$ is the Hamiltonian for this game.

Hamiltonian games can be potential games if they do not have interaction terms. This is because it is a necessary condition that $D_{12} f_1(x) = -D_{12} f_2(x)$.

Example 4.20. Consider the game defined by

$$f(x, y) = x^2 + y^2$$

The potential function is

$$\phi(x, y) = x^2 - y^2$$

Not all Hamiltonian games have to be zero sum.

Example 4.21. Consider the game

$$f_1(x, y) = x(y - c_1), \;\; f_2(x, y) = -(x - c_2) y$$

for fixed constants c_1, c_2. What is happening here is that there is some non-strategic component of the game that is playing a role here. Meaning that if you were to define the game as $(f_1, -f_1)$ then the component of player two's cost $-c_1 x$ would not be something in their control.

4.3.2 Convex Games

We have already previously noted that in the case of non-convex cost functions, we often appeal to local or approximation notions of equilibria.

When the cost functions are convex, however, and the strategy spaces convex, compact (closed and bounded in $\mathbb{R}^n$) we have an important set of results from Rosen 1965. The existence results in this section depend crucially on Kakutani's fixed point theorem. This result extends the celebrated Brouwer fixed point theorem to set value functions, an important feature for games where best response mappings are set valued.

Theorem 4.22. Suppose S is a non-empty, compact, and convex subset of $\mathbb{R}^n$. Let $f : S \to 2^S$ be a set valued function on S with the following properties:

1. f has a closed graph;
2. $f(x)$ is non-empty and convex for all $x \in S$

Then f has a fixed point.

Recall that a μ-strongly convex function h satisfies the following equivalent conditions:

1. $h(y) \geq h(x) + \nabla h(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2, \ \forall x, y$
2. $H(x) = h(x) - \frac{\mu}{2}\|x\|^2$ is convex for all x
3. $(\nabla h(x) - \nabla h(y))^\top (x - y) \geq \mu\|x - y\|^2, \ \forall x, y$
4. $h(\alpha x + (1-\alpha)y) \leq \alpha h(x) + (1-\alpha)h(y) - \frac{\alpha(1-\alpha)}{2}\|x - y\|^2, \ \forall \alpha \in [0, 1]$

A function is convex if

$$h(\alpha x + (1-\alpha)y) \leq \alpha h(x) + (1-\alpha)h(y), \ \forall \alpha \in [0, 1]$$

A. Existence of Pure Nash Equilibria

With this fixed point theorem on hand, Rosen showed that concave (resp. convex) games have pure Nash equilibria. First, let us describe the game set-up.

Consider a game $\mathcal{G} = (f_1, f_2)$—two players for simplicity—defined by cost functions $f_i \in C^2(X, \mathbb{R})$ such that each $f_i(\cdot, x_{-i})$ is convex in x_i and each X_i has the following structure:

$$X_i = \{x_i | \ h_i(x_i) \leq 0\}$$

where $h_i(x_i)$ is a vector valued function having entries $h_{ij}(x_i)$ that are concave functions for each $j = 1, \ldots, k_i$.

As before, a point x^* is a Nash equilibrium of the game if each x_i^* is a best response to x_{-i}^*. There is a slight difference here in that players have constraints so that

$$x_i^* \in \arg\min_{x_i \in X_i} f_i(x_i, x_{-i}^*) = \arg\min_{x_i}\{f_i(x_i, x_{-i}^*) | \ h_i(x_i) \leq 0\} \tag{4.14}$$

Note that Rosen treats the coupled constraints case (meaning that h_i may depend on x_{-i}), but we will not address this here.[3]

Example 4.23. Recall the Cournot game from previous lectures. This is an example of a concave game with constraints.

Consider a scenario in which there are two firms. They simultaneously choose any quantity $q_1, q_2 \geq 0$. The price in the market is given by $p(q_1, q_2) = A - q_1 - q_2$ for some constant A. The firms payoff functions are given by

$$u_i(q_i, q_{-i}) = p(q_1, q_2)q_i - c_i q_{-i}$$

for $i = 1, 2$. Consider the optimization problem that each player faces:

$$\max_{q_i}\{u_i(q_i, q_{-i}) | \ q_i \geq 0\} \tag{4.15}$$

[3]This leads to a more complex equilibrium notion referred to as the generalized Nash equilibrium concept.

Other examples include bimatrix games are convex games on convex strategy spaces (i.e., the simplex) and Bertrand competition.

Define

$$F(x,y) = \sum_{i=1}^{2} f_i(y_i, x_{-i}) \tag{4.16}$$

Definition 4.24. The correspondence $\phi : X \rightrightarrows Y$ is upper semicontinuous if, whenever x_n is a sequence in X converging to a point $x \in X$ and $y_n \in \phi(x_n)$ for all n and $y_n \to y \in Y$, then $y \in \phi(y)$.

Lemma 4.25. The correspondence $\phi : X \rightrightarrows Y$ upper semicontinuous if and only if the graph of ϕ is (relatively) closed in $X \times Y$.

This lemma is a direct consequence of the limit definition of a closed set. Recall that a function $f : X \to \mathbb{R}$ has a closed graph if its graph—i.e. $\mathrm{gr}(f) = \{(x, f(x)) : \ x \in X\}$—is a closed subset of the product space $X \times \mathbb{R}$. A closed set is a set that contains all its limit points.

Theorem 4.26. Consider a game $\mathcal{G} = (f_1, \ldots, f_N)$ where for each $x_{-i} \in X_{-i}$, the functions $f_i(\cdot, x_{-i})$ are convex in x_i and where each X_i is convex, compact subset of $\mathbb{R}^{m_i}$. A Nash equilibrium always exists.

Proof. Consider the set-valued mapping defined by

$$\Phi(x) = \{y|\ F(y,x) = \min_{z \in X} F(z,y)\}$$

Since $F(x,z)$ is continuous and convex in z for a fixed x, Φ is an upper semicontinuous mapping that maps each point of the convex compact set X into a closed convex subset of X. Thus, by Theorem 4.22, there exists a point $x^* \in X$ such that $x^* \in \Phi(x^*)$—i.e.,

$$F(x^*, x^*) = \min_{z \in X} F(x^*, z) \tag{4.17}$$

Moreover, x^* clearly satisfies (4.14) since

$$F(x^*, x^*) = \min_{(z_1,z_2) \in X_1 \times X_2} f_1(z_1, x_2^*) + f_2(x_1^*, z_2) = \min_{z_1 \in X_1} f_1(z_1, x_2^*) + \min_{z_2 \in X_2} f_2(x_1^*, z_2)$$

which means

$$x_i^* \in \arg\min_{x_i \in X_i} f_i(x_i, x_{-i}^*),\ \ i = 1, 2$$

Indeed, suppose that x^* was not a Nash equilibrium. Then, for some player, say $i = 1$, there would be a point $x_1 = \bar{x}_1$ such that $\bar{x} = (\bar{x}_1, x_2^*) \in X$ and $f_1(\bar{x}) < f_1(x^*)$. But then we would have $F(\bar{x}, x^*) > F(x^*, x^*)$ which is clearly contradictory to (4.17). ∎

This result was also independently shown in 1952 by Debreu [16], Fan [22], and Glicksberg [28].

Theorem 4.27 (Debreu, Glicksberg, Fan '52)**.** Consider a continuous game $\mathcal{G} = (f_1, \ldots, f_N)$ such that for each $i \in \mathcal{N}$

- X_i is compact and convex
- $f_i(x_i, x_{-i})$ is continuous in x_{-i}
- $f_i(x_i, x_{-i})$ is continuous and convex in x_i

Then, a pure strategy Nash equilibrium exists.

Proof. Define the best response correspondence for player i as follows:

$$\mathrm{BR}_i(x_{-i}) = \{x_i' \in X_i |\ f_i(x_i', x_{-i}) \leq f_i(x_i, x_{-i}),\ \forall x_i \in X_i\}$$

Define the set of best response correspondences as

$$\mathrm{BR}(x) = \{\mathrm{BR}_i(x_{-i})\}_{i \in \mathcal{N}}$$

Now, we claim we can apply Kakutani's fixed point theorem (Theorem 4.22). To do that we need to show that $\mathrm{BR}(x)$ is non-empty, a convex-valued correspondence, and has a closed graph.

First, $\mathrm{BR}(x)$ is non-empty. To see this, consider

$$\mathrm{BR}_i(x_{-i}) = \arg\min_{z \in X_i} f_i(z, x_{-i})$$

where X_i is non-empty and compact, and f_i is continuous in z by assumption. Then, by Weirstrass's theorem $\mathrm{BR}(x)$ is non-empty.

Now, $\mathrm{BR}(x)$ is a convex-valued correspondence. This is due to the fact that $f_i(x_i, x_{-i})$ is convex (or quasi-convex is enough) in x_i. Indeed, suppose not. Then, there exists some i and some $x_{-i} \in X_{-i}$ such that $\mathrm{BR}_i(x_{-i}) \in \arg\min_{z \in X_i} f_i(z, x_{-i})$ is not convex. This implies that there exists $z, y \in \in \mathrm{BR}_i(x_{-i})$ while $\lambda z + (1-\lambda) y \notin \mathrm{BR}_i(x_{-i})$. That is,

$$\lambda f_i(z, x_{-i}) + (1-\lambda) f_i(y, x_{-i}) < f_i(\lambda z + (1-\lambda) y)$$

which contradicts the convexity of f_i in x_i. Thus, $\mathrm{BR}(x)$ is convex-valued.

■

Glicksberg showed in 1952 [28] that for compact strategy spaces and continuous utility functions there exists *mixed strategy equilibrium* for the game. Dasgupta and Maskin extended this result in 1986 as well to discontinuous games where the utilities are not even quasi-concave!

Example 4.28. (Relaxing Quasi-continuity.) Consider the location game where players choose a point on the circle and their payoffs are

$$f_1(s_1, s_2) = -f_2(s_1, s_2) = \|s_1 - s_2\|_2^2$$

Then there is no pure strategy Nash, yet there is a mixed Nash in this game where both players mix uniformly on the circle.

B. Uniqueness of Pure Nash Equlibria

Rosen additionally showed that under reasonable conditions, we also have uniqueness of the equilibrium.

Assumption 1. The set X contains a point that is strictly interior to every nonlinear—i.e., $\exists \bar{x} \in X$, such that $h_i(\bar{x}) < 0$ for every nonlinear constraint $h_i(x) \geq 0$.

This is a sufficient condition for the satisfaction of the Kuhn-Tucker constraint qualification [6].

The KKT conditions for

$$f_i(x^*) = \min_{x_i}\{f_i(x_i, x^*_{-i})|\ x_i \in X_i\}$$

are given as follows:

$$h_i(x_i^*) \le 0,\ \forall i = 1, 2$$

and for each $i = 1, 2$, there exists $\mu_i \ge 0$, $\mu_i \in \mathbb{R}^{k_i}$ such that

$$\mu_i^\top h_i(x^*) = 0,$$

and

$$f_i(x^*) \le f_i(x_i, x^*_{-i}) + \mu_i^\top h_i(x_i, x^*_{-i}),\ i = 1, 2$$

Since f_i and h_i are convex and differentiable, the above inequality is equivalent to

$$\nabla_i f_i(x^*) + \sum_{j=1}^{k_i} \mu_{ij}\nabla_i h_{ij}(x_i^*) = 0,\ i = 1, 2$$

Now, since $h_{ij}(x_i)$ is convex, we have that for every $z, y \in X$,

$$h_{ij}(z_i) - h_{ij}(y_i) \le (z_i - y_i)^\top \nabla h_{ij}(y_i)$$

Consider the weighted non-negative vector $r \in \mathbb{R}^2$. For each fixed r, define

$$g(x, r) = \begin{bmatrix} r_1\nabla_1 f_1(x) \\ r_2\nabla_2 f_2(x) \end{bmatrix}$$

We call this the *pseudo-gradient* of

$$F(x, r) = r_1 f_1(x) + r_2 f_2(x)$$

Definition 4.29. The function $F(x, r)$ is diagonally strictly convex for $x \in X$ and $r \ge 9$ if for every $z, y \in X$, we have

$$(y - z)^\top g(z, r) + (z - y)^\top g(y, r) < 0$$

With this condition, we can actually show that the equilibrium is unique.

Theorem 4.30. Consider the continuous convex game $\mathcal{G} = (f_1, f_2)$ defined on compact, convex strategy space. If $g(x, r)$ is diagonally strictly concave for some $r = \bar{r} > 0$, then the equilibrium point x^* is unique.

The proof of is essentially argued by contradiction using the KKT conditions for each player's best response optimization problem.

A sufficient condition for the above property is that

$$G(x, r) + G^\top(x, r) > 0$$

for $x \in X$ where $G(x, r)$ is the Jacobian of $g(x, r)$ with respect to x. This can be shown simply using a line integral of the diagonally strictly convex condition stated in Definition 4.29.

Now recall the game Jacobain in the unconstrained case $J(x)$. If an equilibrium x^* exists for a convex game on unconstrained Euclidean spaces $\mathbb{R}^{m_i}$, then $J(x^*)^\top + J(x^*) > 0$ is sufficient to guarantee that it is unique. This is an interesting condition and we will actually see connections to stability of such equilibria under gradient play in the next chapter.

Comments on Games with Coupled Constraints. When the players actions are constrained and the constraint sets are coupled (i.e., a player's action space depends on the actions of others), we define the so called generalized Nash equilibrium concept. Essentially in this class of games, player i's action space X_i depends on the actions of others x_{-i}—that is, $x_i \in X_i(x_{-i})$. Typically the problem of defining an equilibria is approach from a variational inequality perspective. We will not cover this class of games and their equilibria except to mention that people have studied this setting largely from an optimization approach. The flavor of the results are to provide second order sufficient conditions (analogous to constraint qualification in single player nonlinear optimization) for the generalized Nash equilibrium concept. For more detail, please refer to one of the following seminal references on the topic [20, 4].

C. Connections to Monotone Games

Games that satisfy the diagonally strictly convex condition are also know as monotone games and have been a recent subject of study in learning in games (in particular with limited or bandit feedback) since they possess strong properties such as existence and uniqueness of equilibrium (see, e.g., [51, 50, 26, 38]). Indeed, an alternative to the above expression for diagonally strictly convex is the following. Let $g(x) = (D_1 f_1(x), D_2 f_2(x))$.

Theorem 4.31. Consider a convex game $\mathcal{G} = (f_1, f_2)$ that satisfies the so called cost monotonicity condition

$$\langle g(x') - g(x), x' - x\rangle \geq 0, \ \forall x, x' \in X \tag{4.18}$$

with equality if $x = x'$. Then $\mathcal{G}$ admits a unique Nash equilibrium.

Moreover, if (4.18) holds with equality only if $x = x'$ (meaning it is a necessary condition), then the game is called strictly monotone.

This is highlight an equivalence between strictly convex games and strictly monotone games. Indeed, in the above if we let $x'_{-i} = x_{-i}$, then (4.18) gives us that

$$\langle D_i f_i(x'_i, x_{-i}) - D_i f_i(x_i, x_{-i}), x'_i - x_i\rangle \geq 0, \ \forall x_i, x'_i \in X_i, \ x_{-i} \in X_{-i}. \tag{4.19}$$

This expression is equivalent to $f_i(\cdot, x_{-i})$ being strictly convex.

Proposition 4.32. Any game satisfying (4.18) (i.e., any monotone game) is a convex game.

The tools in this line of work on monotone games draw on variational inequalities [21], and more detail will be provided on this body of work in the chapters on learning in games.

Towards understanding convergence of learning dynamics such as follow-the-regularized-leader, the notion of variationally stable equilibrium has been introduced.

Definition 4.33. Consider an equilibrium point x^* for a monotone game $\mathcal{G} = (f_1, f_2)$. We say x^* is variationally stable if there exists a neighborhood U of x^* such that

$$\langle v(x), x - x^*\rangle \geq 0, \ \forall x \in U$$

with equality if and only if $x = x^*$.

It applies in particular to diagonally strictly convex (aka monotone) games, and the sufficient condition on the game Jacobain given by Rosen is a sufficient condition for variationally stable equilibria.

4.3.3 Non-Convex Games

The most general type of game we will consider are games defined by non-convex cost functions. As noted before, analogous to the non-convex single player optimization case, since the cost functions are non-convex we need local notions of equilibria, the definition of which we recall here for ease of access.

Definition 4.34 (Local Nash Equilibrium)**.** A point x^* is a local Nash equilibrium on $U_1 \times U_2$ for the continuous game $(f_1, \dots, f_N)$ if for each $i \in \{1, 2\}$,

$$f_i(x_i^*, x_{-i}^*) \le f_i(x_i, x_{-i}^*) \;\; \forall \; x_i \in U_i.$$

Nash equilibria in non-convex games are NP-hard to compute in general, even in zero-sum settings where there is a lot of structure. Some efforts have been put forth to better understand the complexity in different settings [33, 15]. It is a highly active area of research to define new refinements of Nash equilibria that ensure existence but also are game theoretically meaningful [35, 23, 32]. Motivated by the use of minmax formulations in machine learning and reinforcement learning, the focus over the last couple years has been primarily on zero-sum settings.

Let's focus our attention on zero-sum settings where we will connect minmax problems, including their equilibrium notions, to games.

A. Zero-sum Non-convex Games

Consider the following minmax optimization problem:

$$\min_{x \in X} \max_{y \in Y} f(x, y) \tag{4.20}$$

We can similarly think about the corresponding zero sum game $\mathcal{G} = (f, -f)$ in which the players face the following best response problems:

$$\min_{x \in X} f(x, y) \tag{4.21}$$

$$\max_{y \in Y} f(x, y) \tag{4.22}$$

Recall that in the previous chapter on finite games, we saw the celebrated minmax theorem which states that

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y)$$

This theorem also holds more generally when the cost function f is convex-concave (i.e., convex in x and concave in y). A word of caution however: in the nonconvex setting

$$\min_{x \in X} \max_{y \in Y} f(x, y) \neq \max_{y \in Y} \min_{x \in X} f(x, y)$$

in general.

Because of this, there has been a surge of recent work connecting minmax and maxmin equilibria to Nash equilibria (i.e., points where the maxmin and minmax values are equal). With non-convexities in mind, there are a number of local equilibrium concepts for minmax problems including local Nash.

Stackelberg Formulations of Minmax Problems. In recent work by Fiez et al., the notion of a local Stackelberg equilibrium was introduced in the context of learning in continuous games.

In a *Stackelberg game*, the leader and follower aim to solve the following optimization problems, respectively:

$$\min_{x_1 \in X_1} \{ f_1(x_1, x_2) \big| \; x_2 \in \arg \min_{y \in X_2} f_2(x_1, y) \}, \tag{L}$$

$$\min_{x_2 \in X_2} f_2(x_1, x_2). \tag{F}$$

This contrasts with a *simultaneous* play game in which each player i is faced with the optimization problem $\min_{x_i \in X_i} f_i(x_i, x_{-i})$.

Definition 4.35 (Fiez et al. [23] Local Stackleberg)**.** Consider $U_i \subset X_i$ for each $i \in \{1, 2\}$. The strategy $x_1^* \in U_1$ is a local Stackelberg solution for the leader if, $\forall x_1 \in U_1$,

$$\sup_{x_2 \in \mathcal{R}_{U_2}(x_1^*)} f(x_1^*, x_2) \leq \sup_{x_2 \in \mathcal{R}_{U_2}(x_1)} f(x_1, x_2),$$

where $\mathcal{R}_{U_2}(x_1) = \{ y \in U_2 | - f(x_1, y) \leq -f(x_1, x_2), \forall x_2 \in U_2 \}$. Moreover, (x_1^*, x_2^*) for any $x_2^* \in \mathcal{R}_{U_2}(x_1^*)$ is a local Stackelberg equilibrium on $U_1 \times U_2$.

Note that this definition is not restricted to zero-sum settings. Indeed, replacing $-f$ with f_2 and f with f, the definition holds for the general sum game (f_1, f_2). This is one of the benefits of the Stackelberg formulation versus some of the other notions we consider in this section.

Analogous to the differential Nash notion we saw earlier in the chapter, there is also a differential Stackelberg equilibrium concept.

Definition 4.36 (Differential Stackelberg [23])**.** The joint strategy $x^* = (x_1^*, x_2^*) \in X$ is a differential Stackelberg equilibrium if $Df_1(x^*) = 0$, $D_2 f_2(x^*) = 0$, $D^2 f_1(x^*) > 0$, and $D_2^2 f_2(x^*) > 0$.

Classical Local Minmax Equilibrium Concepts.

Definition 4.37 (Evtushenko [17, 18] Local MinMax)**.** A point (x^*, y^*) is a Evtushenko local minmax equilibrium if

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, \tilde{y}(x))$$

where $\tilde{y}(x) \in \bar{y}(x)$ with

$$\bar{y}(x) = \left\{ \tilde{y} : \; \max_{y \in Y} f(x, y) = f(x, \tilde{y}) \right\}$$

Alternatively, a point is said to be a Evtushenko local minmax equilibrium if $x^* \in \mathcal{X}^*$ and $y^* \in \bar{y}(x)$ where

$$\mathcal{X}^* = \left\{ x_* : \; \min_{x \in X} f(x, \bar{y}(x)) = f(x_*, \bar{y}(x_*)) \right\}$$

An Evtushenko local minmax is related to a local Stackelberg with the difference being that an Evtushenko local minmax defines the best response or reaction set of the max player in a global sense—that is, $\bar{y}(x)$ is the set of points that are global maxima given x. On the other hand, a local Stackelberg considers only the points in a local neighborhood U_2.

Local Minmax with Local Robustness Characteristics. Also closely related to a local Stackelberg is the notion of a local Minmax introduced by Jin et al.

Definition 4.38 (Jin et al. [32] Local MinMax)**.** A point (x^*, y^*) is a local minmax point of f if there exists $\delta_0 > 0$ and a function $h(\delta) \to 0$ as $\delta \to 0$, such that for any $\delta \in (0, \delta_0]$ and any (x, y) satisfying $\|x - x^*\| \leq \delta$ and $\|y - y^*\| \leq \delta$, we have

$$f(x^*, y) \leq f(x^*, y^*) \leq \max_{y' \in B_{h(\delta)}(y^*)} f(x, y')$$

This definition does not require the min player to take into consideration only points on the reaction curve (locally or globally) but rather looks in a neighborhood of the reaction curve. This is an interesting definition—in particular for machine learning applications—because in a sense it provides some robustness to perturbations in x that would in turn lead to perturbations of the best response curve at least locally.

Comments. what All these definitions have some draw backs in that existence becomes challenging when the maximizing player is not (strongly) concave. This is open line of research.

Part II

Learning in Games

Learning is often used as a way to explain how agents arrive at equilibria via a tâtonnement process. In addition learning algorithms are often used to justify equilibrium notions (e.g., greedy max, etc.). Many sorts of adjustment processes, including biological evolution (studied in, e.g., evolutionary game theory), have been said to broadly involve *learning*, In this context, it is often hard to draw a line between learning and other types of adaption. For example, contrast players having knowledge of the current distribution of strategies for opponents, and then playing a best response to it versus each player receiving a a noisy signal of the current distribution, and then forming a best response.

There are several "natural" learning dynamics used in game theory to model this process of agents grappling towards an equilibriated state. In particular, we will study learning rules that have the following properties:

- the learning rule will provide a complete specification of how each player uses observations to change or update their strategies/actions;
- players are self-interested and play according to their interests within the context of the learning rule (as opposed to the learning rule being designed to find an equilibrium);[4]
- the learning rules are uncoupled meaning that each player's update is independent of the payoff functions and rules of the other players;
- we will use performance of the learning rules as a check on their plausibility, not as the goal itself.

In the remainder of this part, we will review some of the classical approaches to learning in games (including those focused largely on finite action spaces or mixed strategies in finite action games), as well as modern approaches to analyzing learning in games. The latter includes learning in continuous action space games, learning with limited/partial or bandit feedback, and dynamical systems theory analysis tools.

[4]The parenthetical here is another line of work that has been studied in applications including machine learning. Specifically, in the context of leveraging a game theoretic abstraction for the purpose of training a robust machine learning algorithm, often the autonomy of the agents is not important and rather, seeking out or computing an equilibrium is the ultimate goal. In this case, then designing the learning rule to find an equilibrium is considered. There are several such examples in the machine learning literature including [36, 47].

Chapter 5

A Review of "Natural" Learning Dynamics

Learning is often used as a way to explain how agents arrive at equilibria via a tâtonnement process. In addition learning algorithms are often used to justify equilibrium notions (e.g., greedy max, etc.). Many sorts of adjustment processes, including biological evolution (studied in, e.g., evolutionary game theory), have been said to broadly involve *learning*, In this context, it is often hard to draw a line between learning and other types of adaption. For example, contrast players having knowledge of the current distribution of strategies for opponents, and then playing a best response to it versus each player receiving a a noisy signal of the current distribution, and then forming a best response.

There are several "natural" learning dynamics used in game theory to model this process of agents grappling towards an equilibriated state. In particular, we will study learning rules that have the following properties:

- the learning rule will provide a complete specification of how each player uses observations to change or update their strategies/actions;
- players are self-interested and play according to their interests within the context of the learning rule (as opposed to the learning rule being designed to find an equilibrium);[1]
- the learning rules are uncoupled meaning that each player's update is independent of the payoff functions and rules of the other players;
- we will use performance of the learning rules as a check on their plausibility, not as the goal itself.

Some questions that motivate the study of the following include the following:

> *Should we expect strategic players do reach an equilibrium of a game, and if so how and how quickly?*

Affirmative answers to these questions are important in that they potentially justify equilibrium analysis. Properties of equilibria (e.g., a near-optimal objective function value) are not obviously

[1] The parenthetical here is another line of work that has been studied in applications including machine learning. Specifically, in the context of leveraging a game theoretic abstraction for the purpose of training a robust machine learning algorithm, often the autonomy of the agents is not important and rather, seeking out or computing an equilibrium is the ultimate goal. In this case, then designing the learning rule to find an equilibrium is considered. There are several such examples in the machine learning literature including [36, 47].

relevant when players fail to find an equilibrium. Yet, proving that natural learning algorithms rapidly converge to an equilibrium certainly lends plausibility to the utility (i.e., predictive power) of the equilibrium concept.

5.1 A Review of Classical Learning Dynamics

There are several well studied learning dynamics:

- best response dynamics
- gradient play (a version of approximate best response)
- (approximate) fictitious play
- replicator dynamics (continuous version of multiplicative weights)

In subsequent sections, we will see that replicator dynamics (multiplicative weights) are a special case of follow-the-regularized-leader. Projected gradient play can also be seen in this more general family of learning dynamics. However, before getting into that line of results, we briefly review these classical methods.

The majority of the classical results apply to games on finite action space and to learning mixed strategies in such games.

For the sake simplicity, we just consider two player games $\mathcal{G} = (f_1, f_2)$ on $X \times Y$.

Best Response Dynamics. The best response dynamics are fairly natural given that we have an equilibrium concept defined in terms of the best response map and the proof of existence we saw argued via a fixed point theorem applied to the best response map. The best response dynamics are given by

$$x_{k+1} = \arg\min_{x \in X} f_1(x_k, y_k) \tag{5.1}$$

$$y_{k+1} = \arg\min_{y \in Y} f_2(x_k, y_k) \tag{5.2}$$

Best response dynamics work well in certain classes of games such as potential games. Recall that in potential games the players are *essentially* optimizing the same function (i.e., the potential function), and as a result we know there exists a pure strategy Nash equilibrium.

Recall that a potential game (finite or otherwise) is a game that admits a potential function $\phi(x)$ such that, for each $i \in \mathcal{N}$,

$$f_i(x_i, x_{-i}) - f(x_i', x_{-i}) = \phi(x_i, x_{-i}) - \phi(x_i', x_{-i}), \ \forall x_i, x_i' \in X_i.$$

Proposition 5.1. In a finite potential game, from an arbitrary initial outcome, best response dynamics converges to a pure Nash equilibrium.

Gradient Play. Gradient play is a natural approximation to the best response dynamics. Indeed, players simply follow the gradient of their individual costs. That is,

$$x_{k+1} = x_k - \gamma_1 D_1 f_1(x_k, y_k) \tag{5.3}$$

$$y_{k+1} = y_k - \gamma_2 D_2 f_2(x_k, y_k) \tag{5.4}$$

Both gradient play and best response are what are known as "uncoupled" dynamics, and Hart and MasCollel showed that such dynamics need not converge to Nash equlibria [30].

It is also common to consider the continuous time gradient dynamics, which are given by

$$\dot{x} = -D_1 f_1(x, y) \tag{5.5}$$

$$\dot{y} = -D_2 f_2(x, y) \tag{5.6}$$

Gradient play is an important class of learning dynamics as it has been widely studied recently for its application in multi-agent reinforcement learning (e.g., multi-agent policy gradient) and in game theoretic abstractions used to train robust machine learning algorithms such as generative adversarial networks and strategic classification.

Fictitious Play. One of the earliest learning rules studied is known as fictitious play (cf. seminal work by Brown '51 [11]), and the idea is that each player plays a best response to the historical average of their opponent. In this class of learning dynamics, players observe the history of play of their opponents. Consider a two player game $\mathcal{G} = (f_1, f_2)$ on $X \times Y$. Suppose they repeatedly choose an action and receive their payoff from an oracle which determines which payoffs agents get based on the game. Let x_t and y_t be the actions of the two players at time t. Then player 1 observes the "history" of player 2:

$$h_1^t = (y_0, \ldots, y_t)$$

and similarly,

$$h_2^t = (x_0, \ldots, x_t)$$

Then, players update as follows:

$$x^{t+1} \in \arg\min_{x \in X} f_1(x, \bar{y}_t)$$

$$y^{t+1} \in \arg\min_{y \in Y} f_2(\bar{x}_t, y)$$

where

$$\bar{x}_t = \frac{1}{t+1}\sum_{k=0}^{t} x_k, \;\; \bar{y}_t = \frac{1}{t+1}\sum_{k=0}^{t} y_k$$

Replicator Dynamics and Multiplicative Weights. The second classical learning algorithm class is knows as replicator dynamics (or its discrete time counterpart, multiplicative weights) [31]. The replicator dynamics are of the form

$$\dot{x}_i = x_i(f_i(x) - \phi(x)) \tag{5.7}$$

where $\phi(x) = \sum_{j=1}^n x_j f_j(x)$. Here, x_i is the proportion of type i in the population of *species* $x = (x_1, \ldots, x_n)$. The function $f_i(x)$ is the fitness of the type i and $\phi(x)$ is the average population fitness. Essentially, the replicator dynamics is a sort of *consensus* dynamics where types are penalized from being far from the average in terms of their level of fitness.

In the context of games, one can think of x_i as the mixed action profile for a player, f_i as a function that evaluates the payoff for a particular action i, and ϕ as the average payoff overall actions. For example, consider a bimatrix game $\mathcal{G} = (A, B)$ on $\mathcal{X} \times \mathcal{Y}$, that is player one choose mixed action profile $x \in \mathcal{X}$ and player two chooses mixed action profile $y \in \mathcal{Y}$ given payoffs $x^\top Ay$ and $x^\top B^\top y$ respectively. Then, the replicator dynamics are given by

$$\dot{x}_i = x_i((Ay)_i - x^\top Ay),\ i \in \{1, \ldots, n\}$$
$$\dot{y}_j = y_j((Bx)_j - y^\top Bx),\ j \in \{1, \ldots, m\}$$

The replicator dynamics have been studied extensively in the context of bimatrix games in the game theory, biology, and even dynamical systems theory communities [48, 31].

The replicator dynamics are closely related to multiplicative weights where the weights are chosen to be exponential (sometimes this is called exponential weights or the Hedge algorithm). A good reference for these types of algorithms more generally is the following [2]. The more general class of dynamics here is known as follow the regularized leader, and in the context of games the following is a good exposition [37].

To see this connection, we consider a transformation to the payoff space:

$$y_i(t) = \int_0^t f_i(x(\tau))d\tau$$

In particular, there is a one-to-one mapping (a diffeomorphism in fact) between y and x:

$$y_i = \left(\ln \frac{x_{i2}}{x_{i1}}, \ldots, \ln \frac{x_{in}}{x_{i1}}\right)$$

so that

$$x_{ij} = \frac{e^{y_{ij}}}{\sum_{\ell=1}^n e^{y_{i\ell}}}$$

One can easily check that $\dot{x}_{ij}$ is precisely the replicator dynamics. Then from this perspective, the continuous time version multiplicative weights is

$$\dot{y}_{ij} = f_{ij}(x)$$
$$x_{ij} = \Lambda_i(y_{ij})$$

where

$$\Lambda_i(y_i) = \frac{\exp(y_{ij})}{\sum_\ell \exp(y_{i\ell})}$$

And the discrete time version is

$$x_{ij}(t+1) = \frac{x_{ij}(t)\exp(\eta_i f_{ij}(x(t)))}{\sum_\ell x_{i\ell}(t)\exp(\eta_i f_{i\ell}(x(t)))} \tag{5.8}$$

5.2 Correlated Equilibria and No-Regret Dynamics

Learning algorithms in general do not converge to Nash equilibria or even local or approximate Nash equilibria. However there are certain refinements of Nash equilibria to which certain types of learning dynamics are known to converge. The most prominent of these refinements is the correlated equilibrium concept, which allows for mixed strategies to be correlated (i.e., not independent) between players. Since this concept applies to mixed strategies we will only discuss it and the associated learning dynamics in the context of finite games (since mixed strategies in continuous games are probability measures on the underlying continuous actions spaces and this requires knowledge of measure theory to properly define).

The method of proving convergence of learning algorithms to correlated equilibria is to bound the *regret* a player accumulates along its learning path from a sequence of actions generated by the learning algorithm. There are several notions of regret for games. This is an active area of research in fact. In particular, it is of interest to formulate dynamic notions of regret. One of the preeminent researchers in this area (Eva Tardos, Cornell) is currently working in this direction [34, 49]. Note: A survey of this area would make a great project.

Before discussing these regret notions and the types of results that exist, let's remind ourselves of the correlated equilibrium notion.

Consider a game $\mathcal{G} = (\mathcal{N}, X, (f_i)_{i\in\mathcal{N}})$ defined on finite actions spaces X_i, $i \in \mathcal{N}$. Mixed strategies, which allow the players to randomize over their pure strategies, lie in the simplex on X_i—i.e.

$$\mathcal{X}_i = \Delta(X_i) = \{x_i \in \mathbb{R}^{n_i} : \ \mathbf{1}^\top x_i = 1, \ x_{ij} \geq 0, \ j = 1, \dots, n_i\}.$$

To aid in the exposition, let us use the notation $s_i \in X_i$ for a pure strategy for player i.

Definition 5.2 (Pure Nash Equilibrium)**.** A strategy profile $(s_1, \dots, s_N)$ is a pure Nash equilibrium (PNE) of $\mathcal{G}$ if no player can decrease their cost via a unilateral deviation: for each $i \in \mathcal{N}$,

$$f_i(s_i, s_{-i}) \leq f_i(s_i', s_{-i}), \quad \forall \ s_i' \in X_i$$

A mixed strategy x_i is a probability distribution over X_i More formally, recall the definition from earlier lectures.

Definition 5.3 (Mixed Nash Equilibrium.)**.** A set $x = (x_1, \dots, x_N)$ of independent probability distributions is a Mixed Nash equilibrium (MNE) of $\mathcal{G}$ if no player can decrease their expected cost under the product distribution $x = x_1 \times \cdots \times x_N$ via a unilateral deviation: for each i,

$$\mathbb{E}_{s\sim x} f_i(s_i, s_{-i}) \leq \mathbb{E}_{s_{-i}\sim x_{-i}} f_i(s_i', x_{-i}), \quad \forall \ s_i' \in X_i$$

Notice that we have written the inequality in a slightly different format, now writing the expectation with respect to the joint (probability) distribution $x \in \Delta(X_1) \times \cdots \times \Delta(X_N)$. This is exactly the same as before, we are just being more clear about the difference between the pure Nash case and the mixed Nash case in which players are allowed to randomize and hence, the equilibrium concept is formulated with respect to their expected utility.

Recall that Nash's theorem says that mixed Nash equilibria always exist [40]. Complexity theory, however, provides compelling evidence for the computational hardness of finding Nash equilibria (see, e.g., [27, 14, 15, 19]). On the other hand, refinements of Nash are much *easier* to compute from a complexity perspective.

5.2.1 Correlated Equilibria

Observe that in a Nash equilibrium, people randomize independently. For games with multiple Nash equilibria, one may want to allow for randomizations between Nash equilibria by some form of communication prior to the play of the game. This perspective naturally leads to the following notion of equilibrium. In the correlated equilibrium concept, introduced by Aumann in 1974 [3], players observe a correlating signal before making their choice. For example, you may think about introducing a traffic light or stop sign in the Chicken game—i.e., two cars approach each other from opposite directions and must choose to stay the course or swerve.

Definition 5.4 (Correlated Equilibrium)**.** A joint probability distribution x over X is a correlated equilibrium (CE) of $\mathcal{G}$ if, for each i,

$$\mathbb{E}_{s\sim x}[f_i(s_i, s_{-i})|\ s_i] \leq \mathbb{E}_{s_{-i}\sim x_{-i}}[f_i(s_i', s_{-i})|\ s_i], \quad \forall\ s_i' \in X_i$$

Note that in the above definition, $s \sim x$ means that the joint strategy s is distributed according to the probability distribution $x \in \Delta(X)$ where $\Delta(X)$ is the simplex in $\mathbb{R}^n$. Further, $s_{-i} \sim x_{-i}$ means that the joint strategy s_{-i}, which excludes player i's strategy s_i, is distributed according to the marginal distribution x_{-i}.

The following is an intuitive interpretation of the correlation mechanism. There is a 'mediator' who draws an outcome s from the publicly known distribution x, and subsequently privately recommend strategy s_i to each player i. Correlated equilibria require the expected payoff from playing the recommended strategy to be greater than or equal to playing any other strategy.

Example 5.5 (Traffic light game with correlated equilibrium that is not a mixed Nash equilibrium.)**.** Consider the following game:

P1\ P2	stop	go
stop	(0,0)	(0,1)
go	(1,0)	(-5,-5)

If the other player is stopping at an intersection, then you would rather go and get on with it. The worst-case scenario, of course, is that both players go at the same time and get into an accident two pure Nash equilibria: $\{$(stop, go), (go, stop)$\}$. define x by uniformly randomizing between these two pure Nash equilibria. This is not a product distribution, and hence, it cannot be a mixed Nash equilibrium. It is however a correlated equilibrium, and to see this consider the following line of reasoning:

- From the perspective of player 1, if the mediator (stop light) recommends go then player 1 knows that player 2 was recommended to stop.
- Assuming player 2 plays the recommended strategy stop, player 1's best response is to follow its recommended strategy go.
- Similarly, when player 1 is told to stop, it assumes that player 2 will go, and under this assumption, stopping is a best response.

In terms of computation, as we have alluded to, Nash equilibria are notoriously hard to compute because they are fixed points. Correlated equilibria, on the other hand, are computationally efficient to find since the set of correlated equilibria are a convex polytope [43]

A correlated equilibrium requires that following the recommended strategy be a best response in the interim stage—i.e., a correlated equilibrium requires that after a profile s is drawn and recommended, playing s_i is a best response for i conditioned on seeing s_i and given everyone else plays s_{-i}. To relax this fact, other refinements have also been introduced in the literature such as the coarse correlated equilibrium, which only requires the recommended strategy to be a best response at the ex-ante stage—i.e., coarse correlated equilibrium requires only that following the suggested strategy s_i when $s \sim x$ is only a best response in expectation before you see s_i.

Definition 5.6 (Coarse Correlated Equilibrium.)**.** A coarse correlated equilibrium (CCE) is a distribution x over actions S such that for every player i,

$$\mathbb{E}_{s\sim x}[f_i(s_i, s_{-i})] \leq \mathbb{E}_{s\sim x}[f_i(s_i', s_{-i})], \quad \forall\ s_i' \in X_i$$

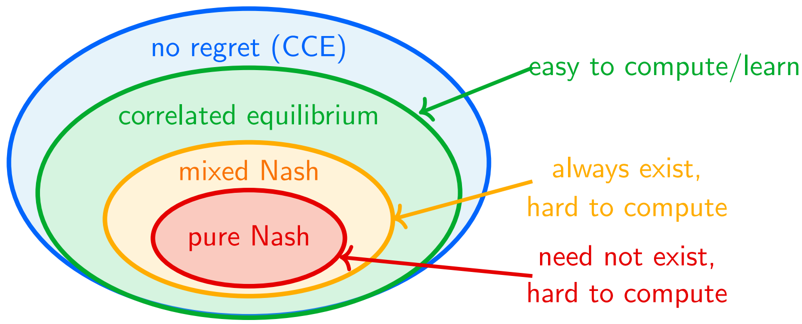

Figure 5.1: Taxonomy of Equilibria and Computability

Note that the set of all distributions satisfying the above is sometimes called the Hannan set.

To see there is a gap between correlated equilibria and coarse correlated equilibria, consider the following example.

Example 5.7. Consider the game defined by

P1\ P2	s_1	s_2	s_3
s_1	(-1,-1)	(1,1)	(0,0)
s_2	(1,1)	(-1,-1)	(0,0)
s_3	(0,0)	(0,0)	(1.1,1.1)

and the strategy $x_i = (1/3, 1/3, 1/3)$ for each $i = 1, 2$. Since the game is symmetric if players play according to this strategy, the payoff (cost) is

$$\tfrac{1}{3} \cdot (-1) + \tfrac{1}{3} \cdot (-1) + \tfrac{1}{3}(1.1) = -0.3$$

Now, consider what would happen if player 1 (without loss of generality) chooses to play the fixed action s_1 or s_2 while the other player continues to randomize. Then the payoff for player 1 would be

$$-\tfrac{1}{3} \cdot 1 + \tfrac{1}{3} \cdot 1 + \tfrac{1}{3} \cdot 0 = 0$$

and the payoff for playing s_3 is strictly positive. This means that given the distribution is a coarse correlated equilibrium since

$$\mathbb{E}_{s \sim x} f_i(s_i, s_{-i}) \le \mathbb{E}_{s \sim x}[f_i(s_i, s_{-i})], \ \forall s_i' \in X_i$$

despite the fact that conditioned on being told to player s_3 is not a best response (because of the positive cost).

5.2.2 Repeated Games

Repeated games consist of a *stage game* that is played for an infinite number of time periods by players who seek to maximize their *discounted* payoffs. Repeated games are one of the more common classes of models examined in economics to model repeated interactions among strategic agents. They give rise to incentives and opportunities that are fundamentally different than those observed in isolated interactions—e.g., the promise of rewards and the threat of punishment in the future of the game can provide incentives for desirable behavior now. Moreover, repeated play gives players (and potentially incentive designers) the opportunity to learn and adapt their strategies.

In what follows, we will distinguish a bit between strategies and actions. Let $\mathcal{G} = \{X, (f_i)_{i\in\mathcal{N}}\}$ denote the stage game. Players repeatedly play $\mathcal{G}$ at times t. Let h_i^t denote the history observed by player i up to time t. Then, H_i^t is the set of all such histories at time t, and $H_i = \cup_t H_i^t$ the set of all histories over time. Then, a policy (or strategy) $\pi_i : H_i \to \mathcal{X}_i$ for player i is a mapping from the the set of histories to a mixed action (or even a pure action depending on the context).[2] In particular, a strategy $\pi = (\pi_1, \ldots, \pi_N)$ induces an sequence of (mixed) action profiles $\{x^t\}_t$. The strategy for player i takes in the history h_t and generates an action x_i^t for that player. That is,

$$\pi_i(h_i^t) = x_i^t \in \mathcal{X}_i$$

Let δ be the common discount factor for all players. The total cost for player i is then

$$F_i(\pi_i, \pi_{-i}) = (1-\delta)\sum_{t=1}^{\infty} \delta^t f_i(x_i^t, x_{-i}^t)$$

Classification of repeated games by available information. What information players have available to them is crucial in determining the outcome of repeated games. There are several types of information structures typically referred to as different types of monitoring:

- **Perfect public monitoring**: At the end of each time period t, players can observe each others' actions and payoffs. With perfect public monitoring $h_i^t = h^t$ for all i and $h^t = (x^1, \ldots, x^{t-1})$.
- **Imperfect public monitoring**: At the end of each time period t, players can only observe a public signal that contains information about the other players' actions, but this signal does not necessarily determine the opponents' actions. That is, we have again that $h_i^t = h^t$ for all i but h^t is not necessarily equal to $(x^1, \ldots, x^{t-1})$.
- **Private monitoring**: At the end of each time period t, players receive a possibly different private signal containing information about what has happened in that round.

Note that one can relate this in the learning dynamics to coupled, uncoupled, and completely uncoupled dynamics where coupling is used to mean information available.

Learning Algorithms as Strategies. Interestingly, we can think of the policy $\pi_i : H_i \to X_i$ as an algorithm, and in the context of the learning in games literature this is precisely what people tend to do. Here, an algorithms $\mathcal{A}_i$ selects a distribution x_i^t

5.2.3 External Regret

To introduce the idea of regret in the context of repeated play games, we will consider algorithms for repeated play of a matrix game with the guarantee that against any opponent, they will perform nearly as well as the best fixed action in hindsight. This is often called the problem of combining expert advice or minimizing *external regret.*

[2] We say mixed more generally here since it encompasses the possibility of pure actions.

To introduce this concept, we will start by taking the view of a single agent against an adversary and hence, we do not add the i subscript here to the action set or actions. Consider a set X of $n \geq 2$ actions. At time each time $t = 1, \ldots, T$, the following occurs:

- A decision maker picks a mixed strategy x^t that is a distribution over its actions X.
- An adversary picks a cost vector $\ell^t : X \to [0, 1]$.
- An actions s^t is chosen according to the distribution x^t and the decision maker incurs a cost $\ell^t(s^t)$.
- The decision maker learns the entire cost vector ℓ^t not just the realized cost $\ell^t(s^t)$.

For example, one may consider X to be different investment strategies or different driving routes between home and work. And, the decision-maker tries to determine the least costly action amongst those in X assuming worst case outcomes. In the multi-player setting (which we will get to), the action set will be the strategy of a single player and the cost vector will be induced by the strategies chosen by all the other players.

Another noteworthy point is that a number of online convex optimization problems can be cast in this decision-maker–adversary framework.

We seek a *good* algorithm for online decision-making problems of the above type. The performance of the algorithm is measured by

$$\text{Regret}(T) = \text{“Total profit of best fixed action in hindsight”} - \text{“Total profit of algorithm”}$$

Definition 5.8 (Time Average Regret)**.** The (time average) regret of the action sequence $s^1, \ldots, s^T$ with respect to a fixed action s is given by

$$\frac{1}{T}\left(\sum_{t=1}^{T} \ell^t(s^t) - \sum_{t=1}^{T} \ell^t(s)\right)$$

Please note there are several other notions of regret including internal and swap regret that are commonly used in analyzing learning in games. We will focus on external regret for now.

Let $\mathcal{A}$ be an online decision-making algorithm. An adversary for $\mathcal{A}$ is a function that takes as input the time index t, the mixed strategies $(x^1, \ldots, x^t)$ produced by $\mathcal{A}$ up to time t, and the realized actions $s^1, \ldots, s^{t-1}$ up to time $t - 1$. It produces as output, the cost vector $\ell^t : S \to [0, 1]$.

Definition 5.9 (No-Regret Algorithm.)**.** An online decision making algorithm $\mathcal{A}$ has *no (external) regret* if the expected regret with respect to every action $s \in X$ is $o(1)$[3] as $T \to \infty$ independent of the adversary—that is,

$$\lim_{T\to\infty} \frac{1}{T}\left(\sum_{t=1}^{T} \ell^t(s^t) - \sum_{t=1}^{T} \ell^t(s)\right) \to 0$$

meaning that the regret grows sublinearly in T.

[3] A function $f \in O(g)$ means that for at least one constant $k > 0$, you can find a constant c such that $0 \leq f(x) \leq kg(x)$ for all $x > c$. Note that $f \in o(g)$ means that the above holds for every choice of $k > 0$. Big-O essentially says that f grows asymptotically no faster than g and Little-o says that the asymptotic growth of f is strictly slower than g. In particular, $f \in o(g)$ means $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$ while $f \in O(g)$ means $\lim_{x\to\infty} \frac{f(x)}{g(x)} < \infty$.

If we think of each action as an "expert" that makes recommendations, then a no-regret algorithms performs asymptotically as well as the best expert. The adversary in the definition of no-regret algorithms is sometimes called an adaptive adversary. As an alternative, an oblivious adversary is the special case in which the cost vector ℓ^t depends only on t (and on the algorithm $\mathcal{A}$).

There are several notable results in the study of no-regret algorithms.

Theorem 5.10 (Informal)**.** The following two results hold for no-regret algorithms.

1. No-regret Algorithms exist and they tend to be simple and have expected regret $O(\sqrt{(\ln n)/T})$ with respect to every fixed action (n is the number of actions).
2. There exists an online decision-making algorithm that, for every $\varepsilon > 0$, has expected regret at most ε with respect to every fixed action after $O\left((\ln n)/\varepsilon^2\right)$ iterations.

The Multiplicative Weights (MW) algorithm (aka Randomized Weighted Majority or Hedge) is a no-regret algorithm. It maintains a weight, intuitively a "credibility," for each action. At each time step the algorithm chooses an action with probability proportional to its current weight. The weight of each action can only decrease, and the rate of decrease depends on the cost of the action.

The main idea comes from the following two formalisms:

1. Past performance of actions should guide which action is chosen now. Since the action choice must be randomized, the probability of choosing an action should be increasing in its past performance (i.e., decreasing in its cumulative cost).
2. Many instantiations of the above idea yield no-regret algorithms. For optimal regret bounds, however, it is important to aggressively punish bad actions—i.e., when an action previously observed to be 'good' turns out to be 'bad', the probability with which it is played should decrease at an exponential rate.

Algorithm 1 Multiplicative Weights Sketch

procedure MW(T) ▷ final time
  $w^1(s) \leftarrow 1, \ \forall s \in X$
  for $t = 1, \ldots, T$ **do**
    $x^t = w^t/\Gamma^t$ where $\Gamma^t = \sum_{s \in X} w^t(s)$ ▷ compute current distribution
    Play an action $s \sim x^t =$ ▷ choose action and play
    Receive loss vector ℓ^t
    For each $s \in X$, $w^{t+1}(s) \leftarrow w^t(s)(1-\varepsilon)^{\ell^t(s)}$ ▷ update weights

In the above algorithm, e.g., if all losses are either 0 or 1 (a typical assumption), then the weight of each action either stays the same (if $\ell^t(s) = 0$) or gets multiplied by $(1-\varepsilon)$ (if $\ell^t(s) = 1$).

As $\varepsilon \to 0$, x^t tends towards the uniform distribution and thus, small values of ε encourage exploration. As $\varepsilon \to 1$, x^t tends to the distribution that puts all its mass on the action that has incurred the smallest cumulative cost so far and thus, large values of ε encourage "exploitation."

5.2.4 No-Regret Dynamics in Games

We now go from single player with adversary to multi-player settings and do so in the cost minimization framework here.

For each time $t = 1, 2, \ldots, T$, the following events occur:

1. Each player i simultaneously and independently chooses a mixed strategy x_i^t using a no-regret algorithm.
2. Each player i receives a cost vector ℓ_i^t where $\ell_i^t(s_i)$ is the expected cost of action s_i when the other players play their chosen mixed strategies—i.e.,

$$\ell_i^t(s_i) = \mathbb{E}_{s_{-i}\sim x_{-i}}[f_i(s_i, s_{-i})]$$

where $x_{-i} = \prod_{j\neq i} x_j$ and f_i is the cost function for agent i.

No-regret dynamics are well defined because no-regret algorithms exist (we did not prove this but we stated it earlier). It is very interesting to note that players *do not have to use the same no-regret algorithm*; they just have to use *some* no-regret algorithm. In addition, while above we have described the players moving simultaneously, the results also extend to players that move sequentially.

No-regret dynamics can be implemented efficiently. Indeed, consider multiplicative weights where n is the (maximum) size of a player's strategy space (across players). Suppose each player uses the multiplicative weights algorithm. Then in each iteration each player does a simple update of one weight per strategy, and only $O((\ln n)/\varepsilon^2)$ iterations of this are required before every player has expected regret at most ε for every strategy. This follows from the analysis in the single player setting.

Perhaps most interestingly we can guarantee that no-regret dynamics converge to coarse correlated equilibria.

Proposition 5.11 (Convergence to CCE.)**.** Suppose after T iterations of no-regret dynamics, every player of a cost minimization game has regret at most ε for each of its strategies. Let $x^t = \prod_{i=1}^N x_i^t$ denote the distribution at time t and let $x = \frac{1}{T}\sum_{t=1}^T x^t$ be the time averaged history of these distributions. Then x is an ε–approximate coarse correlated equilibrium in the sense that for each i,

$$\mathbb{E}_{s\sim x}[f_i(s)] \leq \mathbb{E}_{s_{-i}\sim x_{-i}}[f_i(s_i', s_{-i})] + \varepsilon \quad \forall\ s_i' \in X_i$$

Proof. By definition, for each i

$$\mathbb{E}_{s\sim x}[f_i(a)] = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_{s\sim x^t}[f_i(a)] \tag{5.9}$$

and

$$\mathbb{E}_{s_{-i}\sim x_{-i}}[f_i(s_i', s_{-i})] = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_{s\sim x^t}[f_i(s_i', s_{-i})] \tag{5.10}$$

(since x is the time average distribution). The right hand side of (5.9) and are the time averaged expected costs of player i when playing according to its no-regret algorithm, and the right hand side of (5.10) is the time averaged expected cost when playing the fixed actions s_i' every time instance. By assumption every player has at most regret ε so that (5.9) is at most ε more than (5.10). ■

It is interesting to note that very simple learning algorithms lead remarkably quickly to (approximate) coarse correlated equilibria. In this sense, coarse correlated equilibria are, perhaps unusually, tractable to compute, and hence unusually plausible as a prediction for player behavior.

5.2.5 Swap Regret and ϵ-Correlated Equilibria

The aim for the external regret setting is to design an online algorithm $\mathcal{A}$ that is able to approach the performance of the *best algorithm* from a given class of algorithms say $\mathcal{C}$. Then the goal is to minimize the external regret

$$R_{\mathcal{C}} = L_{\mathcal{A}}(T) - L_{\mathcal{C},\min}(T)$$

where

$$L_{\mathcal{A}} = \sum_{t=1}^{T} \ell_{\mathcal{A}}^t,\ \ell_{\mathcal{A}}^t = \sum_{i=1}^{n} x_i^t \ell_i^t$$

and $L_{\mathcal{C},\min}(T)$ is the minimum regret over the class of algorithms $\mathcal{C}$—i.e., $\min_{A\in\mathcal{C}} L_A(T)$. And, recall that x^t is the distribution across the n actions available to the agent. The class of algorithms $\mathcal{C}$ is the comparison class.

Now, external regret uses a fixed comparison class $\mathcal{C}$, but one can also envision a comparison class that depends on the online algorithm's actions. We can consider modification rules that modify the actions selected by the online algorithm, producing an alternative strategy which we will want to compete against. A modification rule ϕ has as input the history and the current action selected by the online procedure and outputs a, possibly different, action. We denote by ϕ^t the function ϕ at time t, including any dependency on the history.

Effects of Modification Rule. Given a sequence of probability distributions x^t generated by an online algorithms $\mathcal{A}$ and a modification rule ϕ, we define a new sequence of probability distributions $y^t = \phi^t(x^t)$ where

$$y_i^t = \sum_{j:\phi^t(j)=i} x_j^t$$

The loss of the modified sequence is now

$$L_{\mathcal{A},\phi} = \sum_t \sum_i y_i^t \ell_i^t$$

At time t here the modification rule ϕ shifts the probability that $\mathcal{A}$ assigned to action j to action $i = \phi^t(j)$. This means that the modification rule ϕ generates a different distribution as a function of the online algorithm's distribution x^t.

A new notion of regret based on modification rules. Consider the finite set $\mathcal{F}$ of memoryless modification rules (they do not depend on the history).

Definition 5.12. Given a sequence of loss vectors ℓ^t, the regret of $\mathcal{A}$ with respect to the modification rules $\mathcal{F}$ is

$$R_{\mathcal{F}} = \max_{\phi\in\mathcal{F}} \{L_{\mathcal{A}}(T) - L_{\mathcal{A},\phi}(T)\}$$

In this context, external regret is equivalent to letting $\mathcal{F}^{ex}$ be the class of n modification rules ϕ_i where ϕ_i always maps to action i no matter the input. This is because we are comparing to the single best action in hindsight in the case of external regret.

The next natural notion of regret would be to compare to the simple class of policies where you can think back in time and say "ok, what if I was able to switch to some other action? could

I have done better?" This is known as *internal regret* and was introduced by Foster and Vohra in 1998[25]. Specifically, internal regret allows one to modify the online action sequence by changing every occurrence of a given action i to an alternative action j.

Definition 5.13 (Internal Regret)**.** Let $\mathcal{F}^{in}$ be the set of $n(n-1)$ modification rules ϕ_{ij} where $\phi_{ij}(i) = j$ and $\phi_{ij}(i') = i'$ for $i' \neq i$. The internal regret of $\mathcal{A}$ is

$$\max_{\phi \in \mathcal{F}^{in}} \{L_{\mathcal{A}}(T) - L_{\mathcal{A},\phi}(T)\} = \max_{i,j \in X} \left\{ \sum_{t=1}^{T} x_i^t(\ell_i^t - \ell_j^t) \right\}$$

Essentially internal regret compares to the case where at each time you can switch to any other pure action, and compare to this as your benchmark. This is different than comparing to the single best action in hindsight. A good reference on comparing internal and external regret is [8]. The way they explain it is with the problem of purchasing stocks. In particular, we can think of the class of switching policies $\mathcal{F}^{in}$ in the following way.

> *"...every time you bought IBM, you should have bought Microsoft instead."*

Low internal regret algorithms were derived by Hart and Mas-Colell (circa 2000) [29], Foster and Vohra (circa 1997) [24, 25], and Cesa-Bianchi and Lugosi (circa 2003, leverages notion of a potential) [12]. In many of these works, the use of the approachability theorem of Blackwell (1956) has played an important role in some of the algorithms [7]. The Blackwell approachability theorem attempts to answer the following question:

> *Given a matrix M and a set S in n-space, can I guarantee that the center of gravity of the payoffs in a long series of plays is in or arbitrarily near S with probability approaching* 1 *as the number of plays becomes infinite ?*

Here, the center of gravity means the average payoff over a long run of plays.

Blum and Mansour show in [8, 9] that there is a simple online way to efficiently convert any low external regret algorithm into a low internal regret algorithm. This actually gave rise to what is known as swap regret, which allows one to simultaneously swap multiple pairs of actions.

Let $\mathcal{F}^{sw}$ be the class of modification rules which includes all n^n functions $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ where the function ϕ swaps the current online action i with $\phi(i)$ which may be the same or a different action.

Definition 5.14. The swap regret of the algorithm $\mathcal{A}$ is

$$\max_{\phi \in \mathcal{F}^{sw}} \{L_{\mathcal{A}}(T) - L_{\mathcal{A},\phi}(T)\} = \sum_{i=1}^{n} \max_{j \in X} \left\{ \sum_{t=1}^{T} x_i^t(\ell_i^t - \ell_j^t) \right\}$$

Proposition 5.15. Since $\mathcal{F}^{ex} \subset \mathcal{F}^{sw}$ and $\mathcal{F}^{in} \subset \mathcal{F}^{sw}$, both external and internal regret are upper bounded by swap regret.

Note that in swap, internal and external regret, the modification functions do not depend on history. There are extensions of these ideas to history dependent swap rules and policies [8],

One can show that if there are n actions, then the swap regret is bounded by n times the internal regret.

Coming back to games. The importance of internal and swap regret is due to the fact that they are connected to correlated equilibria.

Consider a game $\mathcal{G} = (\mathcal{N}, X, (f_i)_{i\in\mathcal{N}})$. We consider a player i that plays game $\mathcal{G}$ for T time steps using an online algorithm $\mathcal{A}_i$. At time step t, player i plays a distribution (mixed action) x_i^t while the other players play joint distribution x_{-i}^t. Let $\ell_{\mathcal{A}_i}^t$ denote the loss of player i at time t:

$$\ell_{\mathcal{A}_i}^t = \mathbb{E}_{s\sim x^t}[f_i(s_i^t, s_{-i}^t)]$$

and the cumulative loss is

$$L_{\mathcal{A}_i}(T) = \sum_{t=1}^{T} \ell_{\mathcal{A}_i}^t$$

The loss vector at time t for player i is

$$\ell_i^t = (\ell_{i1}^t, \ldots, \ell_{in}^t)$$

where

$$\ell_{ij}^t = \mathbb{E}_{s_{-i}^t\sim x_{-i}^t}[f_i(s_{ij}^t, x_{-i}^t)]$$

For simplicity here we assume all players have n actions and we have dropped the explicit dependence on $\mathcal{A}_i$. Hence, ℓ_{ij}^t is the loss player i would have observed if it had player action s_{ij} at time t. The cumulative loss of action $s_{ij} \in X_i$ of player i is

$$L_{ij}(T) = \sum_{t=1}^{T} \ell_{ij}^t$$

and

$$L_{i,\min}(T) = \min_j L_{ij}(T)$$

Recall that constant sum games are games such that the utilities of players sum to a constant for any joint strategy x. Zero-sum games are a special case where the constant is zero. We know that such games achieve value. For example, in zero sum games, the value of the game is the minmax value.

Proposition 5.16. Let $\mathcal{G}$ be a two-player constant sum game with value (v_1, v_2). Suppose that each player $i \in \{1, 2\}$ plays for T steps using online algorithm $\mathcal{A}_i$ with external regret R_i. Then the average loss $\frac{1}{T}L_{\mathcal{A}_i}(T)$ is at most $v_i + R_i/T$.

Proof. Let y be the mixed strategy corresponding to the observed frequencies of the actions player 2 has played—i.e., $y_j = \sum_{t=1}^{T} x_{2,j}^t/T$ where $x_{2,j}^t$ is the weight player 2 gives to action j at time t. Since the game is constant sum, for any mixed strategy for player 2, player 1 has some action $\alpha \in X_1$ such that

$$\mathbb{E}_{s_2\sim y}[f_1(\alpha, x_2)] \leq v_1$$

This implies, in our setting, that if player 1 has always played action α then its loss would be at most $v_1 T$. Thus $L_{1,\min}(T) \leq L_\alpha(T) \leq v_1 T$. Now since player 1 is playing $\mathcal{A}_1$ with external regret R_1, we have

$$L_{\mathcal{A}_1}(T) \leq L_{1,\min}(T) + R \leq v_1 T + R$$

■

This is to say that if players used an $O(\sqrt{T \log n})$ regret algorithm like MWs then the average loss is at most $v_i + O(\sqrt{(\log N)/T})$.

In a repeated game scenario, if each player uses an action selection algorithm whose regret of this form is sublinear in T, then the empirical distribution of the players actions converges to a correlated equilibrium [29], and in fact, the benefit of a deviation from a correlated equilibrium is bounded exactly by R/T, where R is the largest swap regret of any of the players.

To see this, let's first define relevant modification rules and establish connections between them and equilibria. For $s_1, \alpha, \beta \in X_i$ let

$$\text{switch}_i(s_1, \alpha, \beta) = \begin{cases} \beta, & \text{if } s_1 = \alpha \\ s_1 & \text{otherwise} \end{cases}$$

Given a modification function ϕ for player i, we measure the regret of player i with respect to ϕ as the decrease in its loss:

$$\text{regret}_i(s, \phi) = f_i(s) - f_i(\phi(s_i), s_{-i})$$

For instance, if we consider $\phi = \text{switch}(s_1, \alpha, \beta)$ for a fixed $\alpha, \beta \in X_i$, then $\text{regret}_i(s, \phi)$ is a measure of the regret player i has for playing action α rather than β when others player s_{-i}.

Now, a correlated equilibrium, which we have previously introduced independent of regret, is a distribution x over the joint action space with the following property. A correlating device draws a vector of actions $s \in X$ using distribution x over X, and gives player i the action s_i from s. The distribution x is a correlated equilibrium if for each player it is a best response to play the suggested action, provided that the other players do not deviate. In terms of regret, we have the following definition.

Definition 5.17 (Regret definition of CE)**.** For a general-sum game of any finite number of players, a distribution x over the joint action space is a correlated equilibrium if every player would have zero swap regret when playing it. That is, x is a correlated equilibrium if for every player i and actions $\alpha, \beta \in X_i$, we have

$$\mathbb{E}_{s\sim x}[\text{regret}_i(s, \underbrace{\text{switch}_i(\cdot, \alpha, \beta)}_{\phi})] \leq 0$$

Equivalently, we say x is a correlated equilibrium if and only if for any function $\phi : X_i \to X_i$ we have

$$\mathbb{E}_{s\sim x}[\text{regret}_i(s, \phi)] \leq 0$$

Definition 5.18. A joint distribution x over X is an ε-correlated equilibrium if for every player i and for any function $\phi_i : X_i \to X_i$, we have

$$\mathbb{E}_{s\sim x}[\text{regret}_i(s, \phi_i)] \leq \varepsilon$$

Finally, we can get to the main result. The following theorem relates the empirical distribution of the actions performed by each player, their swap regret, and the distance to correlated equilibrium.

Theorem 5.19. Let $\mathcal{G}$ be a game and suppose that for T steps every player follows a strategy that has swap regret at most R. Then, the empirical distribution $\hat{x}$ of the joint actions player by the players is an (R/T)-correlated equilibrium.

Proof. The empirical distribution $\hat{x}$ assigns to every x^t a probability of $1/T$. Fix a function $\phi_i : X_i \to X_i$ for player i. Since player i has swap regret at most R, we have that

$$L_{\mathcal{A}_i}(T) \leq L_{\mathcal{A}_i,\phi_i}(T) + R$$

where $L_{\mathcal{A}_i}(T)$ is the cumulative loss of player i. By the definition of the regret function, we have

$$\begin{aligned} L_{\mathcal{A}_i}(T) - L_{\mathcal{A}_i,\phi_i}(T) &= \sum_{t=1}^{T} \mathbb{E}_{s^t \sim x^t}[f_i(s^t)] - \sum_{t=1}^{T} \mathbb{E}_{s^t \sim x^t}[f_i(\phi_i(s_i^t), s_{-i}^t)] \\ &= \sum_{t=1}^{T} \mathbb{E}_{s^t \sim x^t}[\text{regret}_i(s^t, \phi_i)] \\ &= T\mathbb{E}_{s \sim \hat{x}}[\text{regret}_i(s, \phi_i)], \text{ since } \hat{x} = \tfrac{1}{T} x^t \end{aligned}$$

Thus, for any $\phi_i : X_i \to X_i$, we have

$$\mathbb{E}_{s \sim \hat{x}}[\text{regret}_i(s, \phi_i)] \leq \frac{R}{T}$$

■

This results says that the payoff of each player is its payoff in some approximate correlated equilibrium. In addition, it relates the swap regret to the distance from equilibrium. Note that if the average swap regret vanishes then the procedure converges, in the limit, to the set of correlated equilibria.

5.3 Follow The Regularized Leader

We have not discussed how regret of algorithms connects to equilibrium concepts, but we have really only mentioned one algorithm, namely Multiplicative Weights. The multiplicative weights algorithm is simple to define and analyze, and it has several applications, but perhaps it is not clear where this algorithm comes from. We saw briefly the connection to continuous time replicator dynamics. In fact it is connected to a more general class of learning dynamics known as follow the regularized leader.

The solution x^t proposed by the algorithm at time t can only depend on the previous loss function $\ell^1, \ldots, \ell^{t-1}$. But, the question is *how should it depend on it?* If the offline optimal solution x is consistently better than all others at each time step, then we would like x^t to be that solution, so we want x^t to be a solution that would have worked well in the previous steps. The most extreme way of implementing this idea is what is know as *Follow the Leader* algorithm (abbreviated FTL), in which we set the solution at time t to be

$$x^t = \arg\min_{x \in \mathcal{X}} \sum_{k=1}^{t-1} \ell^k(x)$$

This algorithm can perform poorly, since it can easily overfit to the past data which itself may be bad. The way we fix this (and a common theme more generally in machine learning if we want to avoid overfitting and promote generalization) is to regularize. That is, we do not want the solution to change much from round to round.

Consider the FTRL (follow the regularized leader) update

$$x^t = \arg\min_{x \in \mathcal{X}} r(x) + \sum_{k=1}^{t-1} \ell^k(x)$$

Typically, the function $r(\cdot)$ is chosen to be strictly convex and to take values that are rather big in magnitude.

Now consider a game setting in which each player is updating according to the above follow the regularized leader dynamics. These dynamics actually have an equivalent continuous time formulation which has been recently studied in adversarial learning (and more generallly zero sum games) [37]:

$$y_i(t) = y_i(0) + \int_0^t v_i(x(\tau))\, d\tau \tag{5.11}$$

$$x_i(t) = Q_i(y_i(t)) \tag{5.12}$$

where

$$u_i(x) = \langle v_i(x), x_i \rangle = \sum_{\alpha_i \in X_i} x_{i,\alpha_i} v_{i,\alpha_i}(x)$$

with

$$v_i(x) = (v_{i,\alpha_i}(x))_{\alpha_i \in X_i}$$

being the vector of payoffs to player i given they player strategy α_i and the remaining players play mixed strategy x_{-i}. In addition, $Q_i : \mathbb{R}^{|X_i|} \to \mathcal{X}_i$ is the choice map defined by

$$Q_i(y_i) = \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - r_i(x_i)\}$$

Suppose that r is continuous and strictly convex on $\mathcal{X}_i$ (for each i) and is smooth on the relative interior of every face of $\mathcal{X}_i$ including $\mathcal{X}_i$ itself. The choice map with the regularizer essentially acts to smooth the hard argmax correspondence $y_i \mapsto \arg\max_{x_i \in \mathcal{X}_i} \langle y_i, x_i \rangle$.

As already alluded two the two quintessential examples of FTRL dynamics are multiplicative weights (which we get by using the entroptic regularizer $r_i(x) = \sum_{\alpha_i \in X_i} x_{i,\alpha_i} \log x_{i\alpha_i}$), and the classical Euclidean projection dynamics which are induced by the regularizer $r_i(x) = \frac{1}{2}\|x_i\|^2$.

Theorem 5.20. A player using FTRL has a $O(1/T)$ regret bound independent of what other players do. That is, if a player $i \in \mathcal{N}$ follows FTRL then for every continuous trajectory $x_{-i}(t)$ of the opponents of player i, we have that

$$\mathrm{Reg}_i(T) \leq \frac{\Omega}{T}$$

where $\Omega = \max r_i - \min r_i$ is a positive constant.

Here the regret for the continuous time case is

$$\mathrm{Reg}_i(T) = \max_{z_i \in \mathcal{X}} \frac{1}{T} \int_0^t (u_i(z_i, x_{-i}(\tau)) - u_i(x(\tau)))\, d\tau$$

Part III

Applications in Learning-Based Systems and Market-Based AI

Chapter 6

Adversarial Learning

Chapter 7

Distributionally Robust Optimization

Chapter 8

Decision-making Using Strategically Generated Data

Chapter 9

Multi-Agent Reinforcement Learning

Part IV

Appendices

Appendix A

Review of Function Properties

A.1 Preliminaries

Given a metric space X with metric ρ, we denote $B_r(x)$ as the r-radius ball around the point x using the metric ρ. For example, if $X = \mathbb{R}^n$ is the n-dimensional Eucldiean space endowed with the usual Euclidean metric—i.e., the 2-norm—then $B_1(0)$ is the usual unit norm ball we are familiar with.

The notation $f \in C^r(X, \mathbb{R})$ to denote that a real-valued function $f(x)$ belongs to the family of r-continuously differentiable (sometimes referred to as r-smooth) on a space X functions. In general, throughout this set of lecture notes, we will take $X \subseteq \mathbb{R}^n$.

A.2 Extrema of Functions and Classification of Critical Points

A very important thing to point out here:

> **There is a difference between critical points of functions and critical points of dynamical systems**.

In this section, we focus on critical points of functions.

For a function $f \in C^2(\mathbb{R}^n, \mathbb{R})$, we define critical points as follows.

Definition A.1. A point x is a critical point of f if $Df(x) = 0$.

Critical points have different classifications. They can be (local) minima, (local) maxima, or (local) saddle points, the latter of which is sometimes referred to as a (local) minimax point. In zero-sum games, e.g., we care about the subset of such saddle points whose defining saddle curvature directions are aligned with the players' actions. These correspond to local minimax points for the game—that is to say that not all saddle points of a function are game theoretically meaningful.

For a given function $f \in C^2(\mathbb{R}^n, \mathbb{R})$, we define a local minima as follows.

Definition A.2. A point $x^* \in U \subset \mathbb{R}^n$ is a local minimum for the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ if

$$f(x^*) \leq f(x), \ \forall x \in U$$

A local maximum is defined analogously. Sometimes the definition of a local extremal point is defined given a metric on $\mathbb{R}^n$. For example, a δ-local minima of f is a point x^* such that $f(x^*) \leq f(x)$ for all $x \in B_\delta(x^*)$. Another way you may encounter a local minima is as follows: a point x^* is a local minima if there exists a $\delta > 0$ such that $f(x^*) \leq f(x)$ for all $x \in B_\delta(x^*)$.

The last important critical point type is a saddle point.

Definition A.3. A critical point x—i.e., $Df(x) = 0$—is a saddle point of f if f does not have a local extremum at x^*.

Here, a local extremum is either a local maximum or a local mimimum.

From calculus, you should remember that second order conditions can be used to determine the existence of a local extremum at a critical point.

Proposition A.4. A critical point x^* of the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is a

1. local maximum if $D^2f(x^*) < 0$—i.e., the matrix $D^2f(x^*)$ is strictly negative definite.
2. local minimum if $D^2f(x^*) > 0$—i.e., the matrix $D^2f(x^*)$ is strictly positive definite.

Note that these are sufficient conditions. There are also necessary conditions.

Proposition A.5. If a critical point x^* of the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is a local minimum, then $D^2f(x^*) \geq 0$. Analogously, if a critical point x^* of the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is a local maximum, then $D^2f(x^*) \leq 0$

It is a generic property of the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ that any local minima x^* (resp. local maxima) satisfies $D^2f(x^*) < 0$ (resp. $D^2f(x^*) > 0$). That is, generically the sufficient conditions of Proposition A.4 are also necessary. This means that for almost all (in a formal mathematical sense) sufficiently smooth functions, the second order test in Proposition A.4 is both necessary and sufficient. Moreover, this is a structurally stable property, meaning that if we have a critical point x^* that is a local extremum, then it persists under smooth perturbations of the function. This being said, there are functions that define important classes of games which do not possess this generic property. That is, they belong to a measure zero set of functions. For instance, bilinear functions such as $x^\top y$ define the class of bilinear games, and the extrema of this type of function does not have the generic property described above.

Returning back to the discussion of saddle points, we also define the strict saddle property.

Definition A.6. A critical point x^* of the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is a strict saddle point of f if it has both a strictly negative eigenvalue and a strictly positive eigenvalue—that is, $\lambda_{\min}(D^2f(x^*)) < 0$ and $\lambda_{\max}(D^2f(x^*)) > 0$ where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalue of their argument, respectively.

A.3 Convexity

Appendix B

Review of Dynamical Systems Theory

B.1 Motivation for the Use of Dynamical Systems Theory

To understand convergence or recurrence properties, learning in games requires the use of dynamical systems theory. This is unlike gradient descent in optimization (a single player game) for which we have a number of tools that are parallel to those in dynamical systems theory.

This is because gradient descent in optimization corresponds to gradient flow. For example, consider the objective of minimizing the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$:

$$\min_x f(x)$$

In general, $f(x)$ is non-convex in x (cf. Appendix A for a review of function properties), meaning that we seek local minma of this function. Towards finding local minima, one natural scheme is to follow the gradient of

Beyond gradient descent, however, even more complex algorithms for single player optimization such as momentum-based or symplectic algorithms used, for example, to speed up convergence often require dynamical systems theory and numerical methods for ordinary differential equations (odes). Hence, we will review some basics of numerical methods for odes, focusing primarily on first order methods.

B.2 Review of Numerical Methods for Differential Equations

Appendix C

Dynamic Games

Appendix D

Mechanism Design

Bibliography

[1] Ilan Adler. The equivalence of linear programs and zero-sum games. *International Journal of Game Theory*, 42(1):165–177, 2013.

[2] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

[3] Robert J Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[4] Didier Aussel and Joydeep Dutta. Generalized nash equilibrium problem, variational inequality and quasiconvexity. *Operations research letters*, 36(4):461–464, 2008.

[5] Tamer Başar and Geert Jan Olsder. *Dynamic noncooperative game theory*. SIAM, 1998.

[6] Dimitri P Bertsekas, WW Hager, and OL Mangasarian. *Nonlinear programming*. Athena Scientific Belmont, MA, 1998.

[7] David Blackwell et al. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[8] Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(6), 2007.

[9] Blum, Avrim and Mansour, Yishay. *Learning, Regret Minimization and Equilibria*. 2007.

[10] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[11] George W Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

[12] Nicolo Cesa-Bianchi and Gábor Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51(3):239–261, 2003.

[13] George Bernard Dantzig. *Linear programming and extensions*, volume 48. Princeton university press, 1998.

[14] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[15] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. *arXiv preprint arXiv:2009.09623*, 2020.

[16] Gerard Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, 38(10):886–893, 1952.

[17] Yu G Evtushenko. Iterative methods for solving minimax problems. *USSR Computational Mathematics and Mathematical Physics*, 14(5):52–63, 1974.

[18] Yu.G. Evtushenko. Some local properties of minimax problems. *USSR Computational Mathematics and Mathematical Physics*, 14(3):129 – 138, 1974.

[19] Alex Fabrikant, Christos Papadimitriou, and Kunal Talwar. The complexity of pure nash equilibria. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 604–612, 2004.

[20] Francisco Facchinei and Christian Kanzow. Generalized nash equilibrium problems. *Annals of Operations Research*, 175(1):177–211, 2010.

[21] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[22] Ky Fan. Fixed-point and minimax theorems in locally convex topological linear spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 38(2):121, 1952.

[23] Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in Stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *International Conference on Machine Learning (ICML)*, pages 3133–3144. PMLR, 2020.

[24] Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

[25] Dean P Foster and Rakesh V Vohra. Asymptotic calibration. *Biometrika*, 85(2):379–390, 1998.

[26] Ian Gemp and Sridhar Mahadevan. Online monotone games. *arXiv preprint arXiv:1710.07328*, 2017.

[27] Itzhak Gilboa and Eitan Zemel. Nash and correlated equilibria: Some complexity considerations. *Games and Economic Behavior*, 1(1):80–93, 1989.

[28] Irving L Glicksberg. A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174, 1952.

[29] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[30] Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to nash equilibrium. *American Economic Review*, 93(5):1830–1836, 2003.

[31] Josef Hofbauer, Karl Sigmund, et al. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[32] Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *International Conference on Machine Learning (ICML)*, pages 4880–4889. PMLR, 2020.

[33] Ker-I Ko and Chih-Long Lin. On the complexity of min-max optimization problems and their approximation. In *Minimax and Applications*, pages 219–239. Springer, 1995.

[34] Thodoris Lykouris, Vasilis Syrgkanis, and Éva Tardos. Learning and efficiency in games with dynamic population. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 120–129. SIAM, 2016.

[35] Oren Mangoubi and Nisheeth K Vishnoi. A second-order equilibrium in nonconvex-nonconcave min-max optimization: Existence and algorithm. *arXiv preprint arXiv:2006.12363*, 2020.

[36] Eric V Mazumdar, Michael I Jordan, and S Shankar Sastry. On finding local Nash equilibria (and only local Nash equilibria) in zero-sum games. *arXiv preprint arXiv:1901.00838*, 2019.

[37] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717. SIAM, 2018.

[38] Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, 2019.

[39] Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

[40] John Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

[41] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[42] Martin J Osborne and Ariel Rubinstein. *A course in game theory.* MIT press, 1994.

[43] Christos H Papadimitriou and Tim Roughgarden. Computing correlated equilibria in multi-player games. *Journal of the ACM (JACM)*, 55(3):1–29, 2008.

[44] TES Raghavan. Zero-sum two-person games. *Handbook of game theory with economic applications*, 2:735–768, 1994.

[45] Lillian J Ratliff, Samuel A Burden, and S Shankar Sastry. Characterization and computation of local Nash equilibria in continuous games. In *Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 917–924, 2013.

[46] Robert W Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[47] Florian Schäfer and Anima Anandkumar. Competitive gradient descent. *arXiv preprint arXiv:1905.12103*, 2019.

[48] Sylvain Sorin. Replicator dynamics: Old and new. *Journal of Dynamics & Games*, 7(4):365, 2020.

[49] Eva Tardos. Learning in games with dynamic population. *Simons Institute Lecture: link to slides*, 2015.

[50] Tatiana Tatarenko and Maryam Kamgarpour. Learning generalized nash equilibria in a class of convex games. *IEEE Transactions on Automatic Control*, 64(4):1426–1439, 2018.

[51] Tatiana Tatarenko and Maryam Kamgarpour. Learning nash equilibria in monotone games. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3104–3109. IEEE, 2019.

[52] Heinrich Von Stackelberg. *Market structure and equilibrium*. Springer Science & Business Media, 2010.