An Introduction to Game Theory

Bruce Hajek

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

December 10, 2018

Contents

Preface

Game theory lives at the intersection of social science and mathematics, and makes significant appearances in economics, computer science, operations research, and other fields. It describes what happens when multiple players interact, with possibly different objectives and with different information upon which they take actions. It offers models and language to discuss such situations, and in some cases it suggests algorithms for either specifying the actions a player might take, of for computing possible outcomes of a game.

Examples of games surround us in everyday life, in engineering design, in business, and politics. Games arise in population dynamics, in which different species of animals interact. The cells of a growing organism compete for resources. Games are at the center of many sports, such as the game between pitcher and batter in baseball. A large distributed resource such as the internet relies on the interaction of thousands of autonomous players for operation and investment.

These notes also touch upon *mechanism design*, which entails the design of a game, usually with the goal of steering the likely outcome of the game in some favorable direction.

Most of the notes are concerned with the branch of game theory involving *noncooperative* games, in which each player has a separate objective, often conflicting with the objectives of other players. Some portion of the course will focus on *cooperative* game theory, which is typically concerned with the problem of how to to divide wealth, such as revenue, a surplus of goods, or resources, among a set of players in a fair way, given the contributions of the players in generating the wealth.

This is the latest version of these notes, written in Fall 2017 and being periodically updated in Fall 2018 in conjunction with the teaching of ECE 586GT Game Theory, at the University of Illinois at Urbana-Champaign. Problem sets and exams with solutions are posted on the course website: `https://courses.engr.illinois.edu/ece586gt/fa2017/` and `https://courses.engr.illinois.edu/ece586/sp2013/`. The author would be grateful for comments, suggestions, and corrections.

–Bruce Hajek

Chapter 1

Introduction to Normal Form Games

The material in this section is basic to many books and courses on game theory. See the monograph [13] for an expanded version with applications to a variety of routing problems in networks.

1.1 Normal form games with finite action sets

Among the simplest games to describe are those involving the simultaneous actions of two players. Each player selects an action, and then each player receives a reward determined by the pair of actions taken by the two players. A *pure strategy* for a player is simply one of the possible actions the player could take, whereas a *mixed strategy* is a probability distribution over the set of pure strategies. There is no theorem that determines the pair of strategies the two players of a given game will select, and no theorem that can determine a probability distribution of joint selections, unless some assumptions are made about the objectives, rationality, and computational capabilities of the players. Instead, the typical outcome of a game theoretic analysis is to produce a set of strategy pairs that are in some sort of equilibrium. The most celebrated notion of equilibrium is due to Nash; a pair of strategies is a Nash equilibrium if whenever one player uses one of the strategies, the strategy for the other player is an optimal response. There are, however, other notions of equilibrium as well. Given these notions of equilibrium we can then investigate some immediate questions, such as: Does a given game have an equilibrium pair? If so, is it unique? How might the players arrive at a given equilibrium pair? Are there any computational obstacles to overcome? How high the payoffs for an equilibrium pair compared to payoffs for other pairs?

A two-player normal form game (also called a strategic form game) is specified by an action space for each player, and a payoff function for each player, such that the payoff is a function of the pair of actions taken by the players. If the action space of each player is finite, then the payoff functions can be specified by matrices. The two payoff matrices can be written in a single matrix, with a pair of numbers for each entry, where the first number is the payoff for the first player, who selects a row of the matrix, and the second number is the payoff of the second player, who selects a column of the matrix. In that way, the players select an entry of the matrix. A rich variety of interactions can be modeled with fairly small action spaces. We shall describe some of the most famous examples. Dozens of others are described on the internet.

Example 1.1 *(Prisoners' dilemma)*

There are many similar variations of the prisoners' dilemma game, but one instance of it is given by the following assumptions. Suppose there are two players who committed a crime and are being held on suspicion of committing the crime, and are separately questioned by an investigator. Each player has two possible actions during questioning:

- *cooperate (C) with the other player, by telling the investigator both players are innocent*
- *don't cooperate (D) with the other player, by telling the investigator the two players committed the crime*

Suppose a player goes free if and only if the other player cooperates, and suppose a player is awarded points according to the following outcomes. A player receives

+1 point for not cooperating (D) with the other player by confessing

+1 point if player goes free, i.e. if the other player cooperates (C)

-1 point if player does not go free, i.e. if the other player doesn't cooperate (D)

For example, if both players cooperate then both players receive one point. If the first player cooperates (C) and the second one doesn't (D), then the payoffs of the players are -1, 2, respectively. The payoffs for all four possible pairs of actions are listed in the following matrix form:

		Player 2	
		C (cooperate)	D
Player 1	C (cooperate)	1,1	-1,2
	D	2,-1	0,0

What actions do you think rational players would pick? To be definite, let's suppose each player cares only about maximizing his/her own payoff and doesn't care about the payoff of the other player.

Some thought shows that action D maximizes the payoff of one player no matter which action the other player selects. We say that D is a dominant strategy *for each player. Therefore (D,D) is a dominant strategy equilibrium, and for that pair of actions, both players get a payoff of 0. Interestingly, if the players could somehow make a binding agreement to cooperate with each other, they could both be better off, receiving a payoff of 1.*

Example 1.2 *(Variation of Prisoners' dilemma)*

Consider the following variation of the game, where we give player 1 the option to commit suicide. What actions do you think rational players would pick? To be definite, let's suppose each player cares only about maximizing his/her own payoff and doesn't care about the payoff of the other player.

		Player 2	
		C (cooperate)	D
	C (cooperate)	1,1	-1,2
Player 1	D	2,-1	0,0
	suicide	-100,1	-100, 0

Some thought shows that action D maximizes the payoff of player 1 no matter which action player 2 selects. So D is still a dominant strategy for player 1. Player 2 does not have a dominant strategy. But player 2 could reason that player 1 will eliminate actions C and suicide, because they are (strictly) dominated. In other words, player 2 could reason that player 1 will select action D. Accordingly, player 2 would also select action D. In this example (D,D) is an equilibrium found by elimination of dominated strategies*.*

You can imagine games in which some dominated strategies of one player are eliminated, which could cause some strategies of the other player to become dominated and those could be eliminated, and that could cause yet more strategies of the first player to be dominated and thus eliminated, and so on. If only one strategy remains for each player, that strategy pair is called an equilibrium under iterated elimination of dominated strategies.

Example 1.3 *(Guess 2/3 of average game) Suppose* n *players each select a number from* $[n] \triangleq \{1, 2, \ldots, 100\}$*. The players that select numbers closest to 2/3 the average of all* n *numbers split the prize money equally. In what sense does this game have an equilibrium? Let's try it out in class.*

ECE586GT August 28, 2018
Guess 2/3 of average game
Guesses

1	
1	
1	
1	
1	
2	
2	
5	
8	
9	
10	
11	
14	
15	co-winner (Amir)
15	co-winner (Roshni)
16	
17	
17	
19	
19	
20	
20	
34	
34	
34	
36	
37	
40	
45	
47	
50	
56	
90	
22.03030303	Average
14.68686869	(2/3)*Average

Example 1.4 *(Vickrey second price auction) Suppose an object, such as a painting, is put up for auction among a group of n players. Suppose the value of the object to player i is v_i, which is known to player i, but not known to any of the other players. Suppose each player i offers a bid, b_i, for the object. In a Vickrey auction, the object is sold to the highest bidder, and the sale price is the second highest bid. In case of a tie for highest bid, the object is sold to one of the highest bidders selected in some arbitrary way (for example, uniformly at random, or to the bidder with the longest hair, etc), and the price is the same as the highest bid (because it is also the second highest bid).*

If player i gets the object and pays p_i for it, then the payoff of that player is $v_i - p_i$. The payoff of any player not buying the object is zero.

In what sense does this game have an equilibrium?

Answer: Bidding truthfully is a weakly dominant strategy for each player. That means, no other strategy of a player ever generates a larger payoff, and for any other strategy, there are possible bids by other players, such that bidding truthfully is strictly better.

We pause from examples to introduce some notation and definitions.

Definition 1.5 *(Normal form, also called strategic form, game) A normal form n player game consists of a triplet $G = (I, (S_i)_{i\in I}, (u_i)_{i]\in I})$ such that:*

- *I is a set of n elements, indexing the players. Typically $I = \{1, \ldots, n\} = [n]$.*
- *S_i is the action space of player i, assumed to be a nonempty set*
- *$S = S_1 \times \cdots \times S_n$ is the set of strategy profiles. A strategy profile can be written as $s = (s_1, \ldots, s_n)$.*
- *$u_i : S \to \mathbb{R}$, such that $u_i(s)$ is the payoff of player i.*

Some important notation ubiquitous to game theory is described next. Given a normal form n-player game $(I, (S_i), (u_i))$, an element $s \in S$ can be written as $s = (s_1, \ldots, s_n)$. If we wish to place emphasis on the i^{th} coordinate of s for some $i \in I$, we write s as (s_i, s_{-i}). Here s_{-i} is s with the i^{th} entry omitted. We use s and (s_i, s_{-i}) interchangeably. Such notation is very often used in connection with payoff functions. The payoff of player i for given actions of all players s can be written as $u_i(s)$. An equivalent expression is $u_i(s_i, s_{-i})$. So, for example, if player i switches action to s_i' and all the other players use the original actions, then s changes to (s_i', s_{-i}).

In some situations it is advantageous for a player to randomize his/her action. If the action space for a player is S_i we write Σ_i for the space of probability distributions over S_i. A mixed strategy for player i is a probability distribution σ_i over S_i. In other words, $\sigma_i \in \Sigma_i$. If $f : S_i \to \mathbb{R}$, the value of f for an action $s_i \in S_i$ is simply $f(s_i)$. If σ_i is a mixed strategy for player i, we often use the notational convention $f(\sigma_i) = E_{\sigma_i}[f]$. In particular, if S_i is a finite set, then σ_i is a probability vector, and $f(\sigma_i) = \sum_{s_i \in S_i} f(s_i)\sigma_i(s_i)$. In a normal form game, if the players are using mixed strategies, we assume that the random choice of pure strategy for each player i is made independently of the choices of other players, using distribution σ_i. Thus, for a mixed strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$, the expected payoff for player i, written $u_i(\sigma)$, or, equivalently, $u_i(\sigma_i, \sigma_{-i})$ denotes expectation of u_i with respect to the product probability distribution $\sigma_1 \otimes \cdots \otimes \sigma_n$.

Definition 1.6 *A strategy profile $(s_1, \ldots, s_n)$ for an n-player normal form game $(I, (S_i), (u_i))$ is a* Nash equilibrium *in pure strategies if for each i and any alternative action s_i',*

$$u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}).$$

A strategy profile of mixed strategies $(\sigma_1, \ldots, \sigma_n)$ *for an* n*-player normal form game* $(I, (S_i), (u_i))$ *is a* Nash equilibrium *in mixed strategies if for each* i *and any alternative mixed strategy* σ_i'*,*

$$u_i(\sigma_i, \sigma_{-i}) \geq u_i(\sigma_i', \sigma_{-i}).$$

We adopt the convention that a pure strategy s_i is also a mixed strategy, because it is equivalent to the probability distribution that places probability mass one at the single point s_i. Pure strategies are considered to be degenerate mixed strategies. *Nondegenerate mixed strategies* are those that don't have all their probability mass on one point. *Completely mixed strategies* are mixed strategies that assign positive probability to each action.

The concept of Nash equilibrium is perhaps the most famous equilibrium concept for game theory, but there are other equilibrium concepts. To mention one other, appearing above for the prisoners' dilemma game, is a dominant strategy equilibrium.

Definition 1.7 *Consider a normal form game* $(I, (S_i)_{i\in I}, (u_i)_{i\in I})$*. Fix a player* i *and let* $s_i, s_i' \in S_i$*.*

(i) Strategy s_i dominates strategy s_i' *(or* s_i' *is dominated by* s_i*) for player* i *if:*

$$u_i(s_i', s_{-i}) < u_i(s_i, s_{-i}) \text{ for all choices of } s_{-i}.$$

Strategy s_i *is a dominant strategy for player* i *if it dominates all other strategies of player* i*.*

(ii) Strategy s_i weakly dominates strategy s_i' *(or* s_i' *is weakly dominated by* s_i*) for player* i *if:*

$$\begin{aligned} &u_i(s_i', s_{-i}) \leq u_i(s_i, s_{-i}) \text{ for all choices of } s_{-i} \text{ and} \\ &u_i(s_i', s_{-i}) < u_i(s_i, s_{-i}) \text{ for some choice of } s_{-i}. \end{aligned} \tag{1.1}$$

Strategy s_i *is a* weakly dominant strategy *for player* i *if it weakly dominates all other strategies of player* i*.*

Definition 1.8 *Consider a strategy profile* $s = (s_1, \ldots, s_n)$ *for an* n*-player normal form game* $(I, (S_i), (u_i))$*.*
(i) The profile s *is a* dominant strategy equilibrium *(in pure strategies) if, for each* i*,* s_i *is a dominant strategy for player* i*.*
(ii) The profile s *is a* weakly dominant strategy equilibrium *(in pure strategies) if, for each* i*,* s_i *is a weakly dominant strategy for player* i*.*

Remark 1.9 *A weaker definition of weak domination would be obtained by dropping the requirement* (1.1)*, and in many instances either definition would work. To be definite, we will stick to the definition that includes* (1.1)*, and leave it to the interested reader to determine when the weaker definition of weak domination would also work.*

Proposition 1.10 *(Iterated elimination of weakly dominated strategies (IEWDS) and Nash equilibrium) Consider a finite game* $(I, (S_i)_{i\in I}, (u_i)_{i\in I})$ *in normal form. Suppose a sequence of games is constructed by iterated elimination of weakly dominated strategies (IEWDS). In other words, given a game in the sequence, some player* i *is chosen with some weakly dominated strategy* s_i *and* S_i *is replaced by* $S_i \backslash s_i$ *to obtain the next game in the sequence, if any. If the final game in the sequence has only one strategy profile, then that strategy profile is a Nash equilibrium for the original game.*

Proof. If a game has only one strategy profile, i.e. if each player has only one possible action, then that strategy profile is trivially a Nash equilibrium. It thus suffices to show that if G' is obtained from G by one step of IEWDS and if a strategy profile $\bar{s}$ is a Nash equilibrium for G', then $\bar{s}$ is also a Nash equilibrium for G. In other words, it suffices to show that elimination of a weakly dominant strategy for some player cannot create a new Nash equilibrium. So fix such games G and G' and let s_i' be the action available to some player i in G that was eliminated to get game G' because some action s_i available in G weakly dominated s_i'. Suppose $\bar{s}$ is a Nash equilibrium for G'. Since any action for any player in G' is available to the same player in G, $\bar{s}$ is also a strategy profile for G, and we need to show it is a Nash equilibrium for G. For $j \neq i$, whether $\bar{s}_j$ is a best response to $\bar{s}_{-j}$ for player j is the same for G or G' – the set of possible responses for player j is the same. Thus, $\bar{s}_j$ must be a best response to $\bar{s}_{-j}$ for game G, if $j \neq i$.

It remains to show that $\bar{s}_i$ is a best response to $\bar{s}_{-i}$ for player i in game G. By definition, $\bar{s}_i$ is at least as good a response for player i as any other response in $S_i \backslash \{s_i'\}$. In particular, $\bar{s}_i$ must be at least as good a response as s_i, which in turn is at least as good as s_i', both for player i. Therefore, $\bar{s}_i$ is a best response to $\bar{s}_{-i}$ in game G. Therefore, $\bar{s}$ is a Nash equilibrium for the game G, as we needed to prove. ∎

Proposition 1.10 is applicable to the guess 2/3 of the average game–see the homework problem about it.

Example 1.11 *(Bach or Stravinksy) or opera vs. football, or battle of the sexes*

This two-player normal form game is expressed by the following matrix of payoffs: Player 1 prefers to go to

		Player 2: B	Player 2: S
Player 1	B	3,2	0,0
	S	0,0	2,3

a Bach concert while player 2 prefers to go to a Stravinsky concert, and both players would much prefer to go to the same concert together. There is no dominant strategy. There are two Nash equilibria: (B,B) and (S,S).

The actions B or S are called pure strategies. A mixed strategy is a probability distribution over the pure strategies. If mixed strategies are considered, we can consider a new game in which each player selects a mixed strategy, and then each player seeks to maximize his/her expected payoff, assuming the actions of the players are selected independently.

For this example, suppose player 1 selects B with probability a and S with probability $1-a$, In other words, player 1 selects probability distribution $(a, 1-a)$. If $a = 0$ or $a = 1$ then the distribution is equivalent to a pure strategy, and is considered to be a degenerate mixed strategy. If $0 < a < 1$ the strategy is a nondegenerate mixed strategy. Suppose player 2 selects probability distribution $(b, 1-b)$. Is there a Nash equilibrium for the expected payoff game such that at least one of the players uses a nondegenerate mixed strategy?

Suppose $((a, 1-a), (b, 1-b))$ is a Nash equilibrium with $0 < a < 1$ and $0 \leq b \leq 1$. The expected reward of player 1 is $3ab + 2(1-a)(1-b)$ which is equal to $2(1-b) + a(5b-2)$. In order for a to be a best response for player 1, it is necessary that $5b - 2 = 0$ or $b = \frac{2}{5}$. If $b = \frac{2}{5}$ then player 1 gets the same payoff for action B or S. This fact is an example of the equalizer principle, which is that a mixed strategy is a best response only if the pure strategies it is randomized over have equal payoffs. By symmetry, in order for $(\frac{2}{5}, \frac{3}{5})$ to be a best response for player 2 (which is a nondegenerate mixed strategy), it must be that $a = \frac{3}{5}$. Therefore, $((\frac{3}{5}, \frac{2}{5}), (\frac{2}{5}, \frac{3}{5}))$ is the unique Nash equilibribum in mixed strategies such that at least one strategy is nondegenerate. The expected payoffs for both players for this equilibrium are $3 \cdot \frac{6}{25} + 2 \cdot \frac{6}{25} = \frac{6}{5}$, which is not very satisfactory for either player compared to the two Nash equilibria in pure strategies.

A satisfactory behavior of the players faced with this game might be for them to agree to both play B or both play S, where the decision is made by the flip of a fair coin that both players can observe. Given the coin *flip, neither player would have incentive to unilaterally deviate from the agreement. The expected payoff of each player would be 2.5. This strategy is an example of the players using common randomness (both players observe the outcome of the coin flip) to randomize among two or more Nash equilibrium points.*

Example 1.12 *(Matching pennies) This is a well known zero sum game. Players 1 and 2 each put a penny under their hand, with either (H) "heads" or (T) "tails" on the top side of the penny, and then they simultaneously remove their hands to reveal the pennies. This is a zero sum game, with player 1 winning if the actions are the same, i.e. HH or TT, and player 2 winning if they are different. The game matrix is shown. Are there any dominant strategies? Nash equilibria in pure strategies? Nash equilibrium in mixed*

		Player 2	
		H	T
Player 1	H	1,-1	-1,1
	T	-1,1	1, -1

strategies?

Example 1.13 *(Rock Scissors Paper) This is a well known zero sum game, similar to matching pennies. Two players simultaneously indicate (R) "rock," (S) "scissors," or (P)"paper." Rock beats scissors, because a rock can bash a pair of scissors. Scissors beats paper, because scissors can cut paper. Paper beats rock, because a paper can wrap a rock. The game matrix is shown. Are there any dominant strategies? Nash*

		Player 2		
		R	S	P
	R	0,0	1,-1	-1,1
Player 1	S	-1,1	0,0	1,-1
	P	1,-1	-1,1	0,0

equilibria in pure strategies? Nash equilibrium in mixed strategies?

Example 1.14 *(Identical interest games) A normal form game* $(I, (S_i)_{i\in I}, (u_i)_{i\in I})$ *is an* identical interest game *(or a coordination game) if* u_i *is the same function for all* i*. In other words, if there is a single function* $u : S \to \mathbb{R}$ *such that* $u_i \equiv u$ *for all* $i \in I$*. In such games, the players would all like to maximize the same function* $u(s)$*. That could require coordination among the players because each player* i *controls only entry* s_i *of the strategy profile vector. A strategy profile* s *is a Nash equilibrium if and only if it is a local maximum of* u *in the sense that for any* $i \in I$ *and any* $s'_i \in S_i$, $u(s'_i, s_{-i}) \le u(s_i, s_{-i})$.

1.2 Cournot model of competition

Often in applications, Nash equilibria can be explicitly identified, so there is no need to invoke an existence theorem. And properties of Nash equilibria can be established that sometimes determine the number of Nash equilibria, including possibly uniqueness. These points are illustrated in the example of the Cournot model of competition. The competition is among firms producing a good; the action of each firm i is to select a quantity s_i to produce. The total supply produced by all firms is $s_{tot} = s_1 + \cdots + s_n$. Suppose that the price per unit of good produced depends on the total supply as $p(s_{tot}) = a - s_{tot}$, where $a > 0$, as pictured in Fig. 1.1(a).

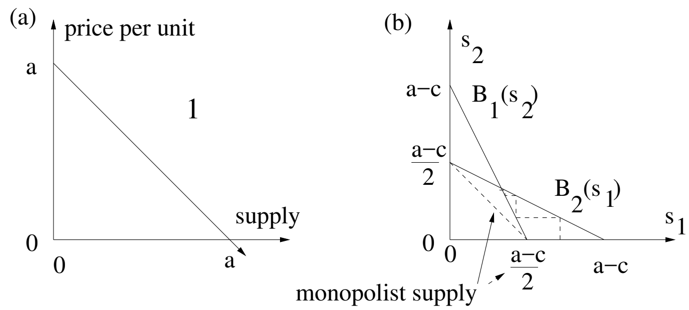


Figure 1.1: (a) Price vs. total supply for Cournot game, (b) Best response functions for two-player Cournot game.

Suppose also that the cost per unit production is a constant c, no matter how much a firm produces. Suppose the set of valid values of s_i for each i is $S_i = [0, \infty)$. To avoid trivialities assume that $a > c$; otherwise the cost of production of any positive amount would exceed the price.

The payoff function of player i is given by $u_i(s) = (a - s_{tot} - c)s_i$. As a function of s_i for s_{-i} fixed, u_i is quadratic and the best response s_i to actions of the other players can be found by setting $\frac{\partial u_i(s)}{\partial s_i} = 0$, or equivalently,

$$(a - s_{tot} - c) - s_i = 0 \quad \text{or} \quad s_i = a - s_{tot} - c.$$

Thus, s_i is the same for all i for any Nash equilibrium, so $s_i = a - ns_i - c$ or $s_i = \frac{a-c}{n+1}$. The total supply at the Nash equilibrium is

$$s_{tot}^{Nash} = \frac{n(a-c)}{n+1},$$

which increases from $\frac{a-c}{2}$ for the case of monopoly (one firm) towards the limit $a - c$ as $n \to \infty$. The payoff of each player i at the Nash equilibrium is thus given by

$$u_i^{Nash} = \left(a - \frac{n(a-c)}{n+1} - c\right)\left(\frac{a-c}{n+1}\right) = \frac{(a-c)^2}{(n+1)^2}.$$

The total sum of payoffs is $nu_i^{Nash} = \frac{n(a-c)^2}{(n+1)^2}$. For example, in the case of a monopoly ($n = 1$) the payoff to the firm is $\frac{(a-c)^2}{4}$ and the supply produced is $\frac{a-c}{2}$. In the case of duopoly (n=2) the sum of revenues is $\frac{2(a-c)^2}{9}$ and the total supply produced is $\frac{2(a-c)}{3}$. In the case of duopoly, if the firms could enter a binding agreement with each other to only produce $\frac{a-c}{4}$ each, so the total production matched the production in the case of monopoly, the firms could increase their payoffs.

To find the best response functions for this game, we solve the equation $s_i = a - s_{tot} - c$ for s_i to get $B_i(s_{-i}) = \left(\frac{a-c-|s_{-i}|}{2}\right)_+$, where $|s_{-i}|$ represents the sum of the supplies produced by the players except player i. Fig. 1.1(b) shows the best response functions B_1 and B_2 for the case of duopoly (n=2). The dashed zig-zag curve starting at $s_1 = \frac{3(a-c)}{4}$ shown in the figure indicates how iterated best response converges to the Nash equilibrium point for two players.

Note: We return to Cournot competition later. With minor restrictions it has an ordinal potential, showing that iterated one-at-a-time best response converges to the Nash equilibrium for the n-player game.

1.3 Correlated equilibria

The concept of correlated equilibrium is due to Aumann [2]. We introduce it in the context of the Dove Hawk game pictured in Figure 1.2. The action D represents cooperative, passive behavior, whereas H represents

		Player 2 D	Player 2 H
Player 1	D	4,4	1,5
	H	5,1	0,0

Figure 1.2: Payoff matrix for the dove-hawk game

aggressive behavior against the play D. There are no dominant strategies. Both (D, H) and (H, D) are pure strategy Nash equilibria. And $((0.5, 0.5), (0.5, 0.5))$ is a Nash equilibrium in mixed strategies with payoff 2.5 to each player. How might the players do better?

One idea is to have the players agree with each other to both play D. Then each would receive payoff 4. However, that would require the players to have some sort of trust relationship. For example, they might enter into a binding contract on the side.

The following notion of correlated equilibrium relies somewhat less on trust. Suppose there is a trusted coordinator that sends a random signal to each player, telling the player what action to take. The players know the joint distribution of what signals are sent, but each player is not told what signal is sent to the other player. For this game, a correlated equilibrium is given by the following joint distribution of signals:

		Player 2 D	Player 2 H
Player 1	D	1/3	1/3
	H	1/3	0

In other words, with probability 1/3, the coordinator tells player 1 to play H and player 2 to play D. With probability 1/3, the coordinator tells player 1 to play D and player 2 to play D. And so on. Both players assume that the coordinator acts as declared. If player 1 is told to play H, then player 1 can deduce that player 2 was told to play D, so it is optimal for player 1 to follow the signal and play H. If player 1 is told to play D, then by Bayes rule, player 1 can reason that the conditional distribution of the signal to player 2 was $(0.5, 0.5)$, so, conditioned on the signal for player 1 being D, the signal to player 2 is equally likely to be D or H. Hence, the conditional expected reward for player 1, given the signal D received, is 2.5 whether player 1 obeys and plays D, or player 1 deviates and plays H. Thus, player 1 has no incentive to deviate from obeying the coordinator. The game and equilibrium are symmetric in the two players, and each has expected payoff $(4+1+5)/3 = 10/3$. This is an example of correlated equilibrium.

A slightly different type of equilibrium, which also involves a coordinator, is to randomize over Nash equilibria, with the coordinator selecting and announcing which Nash equilibria will be implemented. We already mentioned this idea for the Bach or Stravinksy game. For the Dove Hawk game, the coordinator could flip a fair coin and with probability one half declare that the two players should use the Nash equilibrium (H, D) and otherwise declare that the two players should use the Nash equilibrium (D, H). In this case the announcement of the coordinator can be public – both players know what the signal to both players is. Even so, since (H, D) and (D, H) are both Nash equilibria, neither player has incentive to unilaterally deviate from the instructions of the coordinator. In this case, the expected payoff for each player is (5+1)/2 = 3.

Definition 1.15 *A correlated equilibrium for a normal form game $(I, (S_i), (u_i))$ with finite action spaces is a probability distribution p on $S = S_1 \times \cdots \times S_n$ such that for any $s_i, s'_i \in S_i$:*

$$\sum_{s_{-i} \in S_{-i}} p(s_i, s_{-i}) u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} p(s_i, s_{-i}) u_i(s'_i, s_{-i}) \tag{1.2}$$

Why does Definition 1.2 make sense? The interpretation of a correlated equilibrium is that a coordinator randomly generates a set of signals $s = (s_1, \ldots, s_n)$ using the distribution p, and privately tells each player what to play. Dividing each side of (1.2) by $p(s_i)$ (the marginal probability that player i is told to play s_i) yields

$$\sum_{s_{-i} \in S_{-i}} p(s_{-i}|s_i) u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} p(s_{-i}|s_i) u_i(s'_i, s_{-i}). \tag{1.3}$$

Given player i is told to play s_i, the lefthand side of (1.3) is the conditional expected payoff of player i if player i obeys the coordinator. Similarly, given player i is told to play s_i, the righthand side of (1.3) is the conditional expected payoff of player i if player i deviates and plays s'_i instead. Hence, under a correlated equilibrium, no player has an incentive to deviate from the signal the player is privately given by the coordinator.

Remark 1.16 *A somewhat more general notion of correlated equilibria is that each player is given some private information by a coordinator. The information might be less specific than a particular action that the player should take, but the player needs to deduce an action to take based on the information received and on knowledge of the joint distribution of signals to all players, assuming all players rationally respond to their private signals. The actions of the coordinator are modeled by a probability space $(\Omega, \mathcal{F}, p)$ and a set of subsigma algebras $\mathcal{H}_i$ of $\mathcal{F}$, where $\mathcal{H}_i$ represents the information released to player i. Without loss of generality, we can take $\mathcal{F}$ to be the smallest σ-algebra containing $\mathcal{H}_i$ for all i: $\mathcal{F} = \vee_{i \in I} \mathcal{H}_i$, so a correlated equilibrium can be represented by $(\Omega, \{H_i\}_{i \in I}, p, \{S_i\}_{i \in I})$ where each S_i is an $\mathcal{H}_i$ measurable random variable on $(\Omega, \mathcal{F}, p)$, such that the following incentive condition holds for each $i \in I$:*

$$\mathbb{E}\left[u_i(S_i, S_{-i})\right] \geq \mathbb{E}\left[u_i(S'_i, S_{-i})\right] \tag{1.4}$$

for any $\mathcal{H}_i$ measurable random variable S'_i. (Since we've just used S_i for the action taken by a player i we'd need to introduce some alternative notation for the set of possible actions of player i, such as A_i.) In the end, for this more general setup, all that really matters for the expected payoffs is the joint distribution of the random actions, so that any equilibrium in this more general setting maps to an equivalent equilibrium in the sense of Definition 1.15.

A Nash equilibrium in pure strategies is a special case of correlated equilibrium in which the signals are deterministic. A Nash equilibrium in mixed strategies gives a correlated equilibrium in which the signals to different players are mutually independent. Thus, the notion of correlated equilibrium offers more equilibrium possibilities than Nash equilibria, although implementation relies on the availability of a trusted coordinator and private information channels from the coordinator to the players. Geometrically, the set of correlated equilibria is a set of probability vectors satisfying linear constraints, so it is a convex polytope. For finite games it is nonempty because it includes the distributions corresponding to Nash equilibrium in mixed strategies.

1.4 On the existence of a Nash equilibrium

This section describes theorems that ensure the existence of a Nash equilibrium for a normal form game under fairly general conditions. If there is only one player in the game, say player 1, then a Nash equilibrium is simply a strategy $s_1 \in S_1$ that maximizes the function $u_1(s_1)$ over all of S_1. In that case the Weierstrass extreme value theorem (see Section 1.7.1) gives a sufficient condition for the existence of a Nash equilibrium. Fixed point theorems, as described next, give sufficient conditions for existence of Nash equilibrium in games with more than one player.

Theorem 1.17 *(Bauer fixed point theorem) Let S be a simplex*[1] *or unit ball in $\mathbb{R}^n$ and let $f : S \to S$ be a continuous function. There exists $s^* \in S$ such that $f(x^*) = s^*$.*

For a proof see, for example, [6]

Theorem 1.18 *(Kakutani fixed point theorem) Let $f : A \rightrightarrows A$ be a set valued function satisfying the following conditions:*

(i) A is a compact, convex, nonempty subset of $\mathbb{R}^n$ for some $n \geq 1$.

(ii) $f(x) \neq \emptyset$ for $x \in A$.

(iii) $f(x)$ is a convex set for $x \in A$.

(iv) The graph of f, $\{(x, y) : x \in A, y \in f(x)\}$ is closed.

There exists x^* so that $x^* \in f(x^*)$.

Some examples are shown in Figure 1.3 with $A = [0, 1]$.

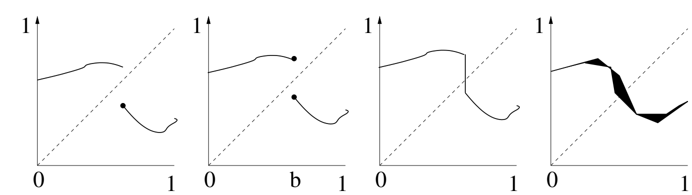


Figure 1.3: Examples of correspondences from $[0, 1]$ to $[0, 1]$ illustrating conditions of Kakutani theorem.

Theorem 1.19 *(Existence of Nash equilibria – Nash) Let $(I, (S_i : i \in I), (u_i : i \in I))$ be a finite game in normal form (so I and the sets S_i all have finite cardinality). Then there exits a Nash equilibrium in mixed strategies.*

Proof. Let Σ_i denote the set of mixed strategies for player i (i.e. the set of probability distributions on S_i). Let $B_i(\sigma_{-i}) \triangleq \arg\max_{\sigma_i} u_i(\sigma_i, \sigma_{-i})$. Let $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$. Let σ denote an element of Σ, so $\sigma = (\sigma_1, \ldots, \sigma_n)$. Let $B(\sigma) = (B_1(\sigma_{-1}) \times \ldots \times B_n(\sigma_{-n}))$. The Nash equilibrium points are the fixed points of B. In other words, σ is a Nash equilibrium if and only if $\sigma \in B(\sigma)$. The proof is completed by invoking the Kakutani fixed point theorem for Σ and B. It remains to check that Σ and B satisfy conditions (i)-(iv) of Theorem 1.18. (i) For each i, the set of probability distributions Σ_i is compact and convex, and so the same is true of

[1] A simplex $S \subset \mathbb{R}^n$ is the convex hull of a finite set of points $x_1, \ldots, x_k \in \mathbb{R}^n$ with the following property. Any $x \in S$ has a unique representation as a convex combination of $x_1, \ldots, x_k$.

the product set Σ. (ii)-(iii) For any $\sigma \in \Sigma$ and $i \in I$, the best response set $B_i(\sigma_{-i})$ is the set of all probability distributions on Σ_i that are supported on the best response actions in S_i, so $B_i(\sigma_{-i})$ is a nonempty convex set for each i, and hence the product set $B(\sigma)$ is also nonempty and convex.

(iv) It remains to verify that the graph of B is closed. So let $(\sigma^{(n)}, \widehat{\sigma}^{(n)})_{n\geq 1}$ denote a sequence of points in the graph (so $\sigma^{(n)} \in \Sigma$ and $\widehat{\sigma}^{(n)} \in B(\sigma^{(n)})$ for each n) that converges to a point $(\sigma^{(\infty)}, \widehat{\sigma}^{(\infty)})$. It must be proved that the limit point is in the graph, meaning that $\widehat{\sigma}^{(\infty)} \in B(\sigma^{(\infty)})$, or equivalently, that $\widehat{\sigma}_i^{(\infty)} \in B_i(\sigma_{-i}^{(\infty)})$ for each $i \in I$. So fix an arbitrary $i \in I$.

Fix an arbitrary strategy $\sigma_i' \in \Sigma_i$. Then

$$u_i(\widehat{\sigma}_i^{(n)}, \sigma_{-i}^{(n)}) \geq u_i(\sigma_i', \sigma_{-i}^{(n)}) \tag{1.5}$$

for all $n \geq 1$. The function $u_i(\sigma_i, \sigma_{-i})$ is an average of payoffs for actions s_j selected independently with distribution σ_j for each j, so it is a continuous function of σ. Therefore, the left and right hand sides of (1.5) converge to $u_i(\widehat{\sigma}_i^{(\infty)}, \sigma_{-i}^{(\infty)})$ and $u_i(\sigma_i', \sigma_{-i}^{(\infty)})$, respectively. Since weak inequality "$\geq$" is a closed relation, (1.5) remains true if the limit is taken on each side of the inequality, so $u_i(\widehat{\sigma}_i^{(\infty)}, \sigma_{-i}^{(\infty)}) \geq u_i(\sigma_i', \sigma_{-i}^{(\infty)})$. Thus, $\widehat{\sigma}_i^{(\infty)} \in B_i(\sigma_{-i}^{(\infty)})$ for each $i \in I$, so $\widehat{\sigma}^{(\infty)} \in B(\sigma^{(\infty)})$. This completes the proof that the graph of B is closed, and hence it also completes the proof of the theorem. ■

Remark 1.20 *Although Nash's theorem guarantees that any finite game has a Nash equilibrium in mixed strategies, the proof is based on a nonconstructive fixed point theorem. It is believed to be a computationally difficult problem to find a Nash equilibrium except for certain classes of games, most importantly, zero sum two-player games. To appreciate the difficulty, we can see that the problem of finding a mixed strategy Nash equilibrium has a combinatorial aspect. A difficult part is to find subsets $S_i^o \subset S_i$ of the action sets S_i of each player such that the support set of σ_i is S_i^o. In other words, $S_i^o = \{a \in S_i : \sigma_i(a) > 0\}$. Let n_i^o be the cardinality of S_i^o. Then a probability distribution on S_i^o has $n_i^o - 1$ degrees of freedom, where the -1 comes from the requirement the probabilities add to one. So the total number of degrees of freedom is $\sum_{i\in I}(n_i^o - 1)$. And part of the requirement for Nash equilibrium is that for each i, $u_i(a, \sigma_{-i})$ must have the same value for all $a \in S_i^o$. That can be expressed in terms of $n_i^o - 1$ equality constraints. Thus, given the sets $(S_i^o)_{i\in I}$, the total degrees of freedom for selecting a Nash equilibrium σ is equal to the total number of equality constraints. In addition, an inequality constraint must be satisfied for each action of each player that is used with zero probability.*

Pure strategy Nash equilibrium for games with continuum strategy sets Suppose $C \subset \mathbb{R}^n$ such that C is nonempty and convex.

Definition 1.21 *A function $f : C \to \mathbb{R}$ is quasi-concave if the t-upper level set of f, $L_f(t) = \{x : f(x) \geq t\}$ is a convex set for all t. In case $n = 1$, a function is quasi-concave if there is a point $c_o \in C$ such that f is nondecreasing for $x \leq c_o$ and nonincreasing for $x \geq c_o$.*

Theorem 1.22 *(Debreu, Glicksberg, Fan theorem for existence of pure strategy Nash equilibrium) Let $(I, (S_i : i \in I), (u_i : i \in I))$ be a game in normal form such that I is a finite set and*

(i) S_i is a nonempty, compact, convex subset of $\mathbb{R}^{n_i}$.

(ii) $u_i(s)$ is continuous on $S = S_1 \times \cdots \times S_n$.

(iii) $u_i(s_i, s_{-i})$ is quasiconcave in s_i for any s_{-i} fixed.

Then the game has a pure strategy Nash equilibrium.

Proof. The proof is similar to the proof of Nash's theorem for finite games, but here we consider best response functions for pure strategies. Thus, define $B : S \rightrightarrows S$ by $B(s) = B_1(s_{-1}) \times \cdots \times B_n(s_{-n})$, where B_i is the best response function in pure strategies for player i: $B_i(s_{-i}) = \arg\max_{a \in S_i} u_i(a, s_{-i})$. It suffices to check that conditions (i)-(iv) of Theorem 1.18 hold. (i) S is a nonempty compact convex subset of $\mathbb{R}^n$ for $n = n_1 + \cdots + n_n$. (ii) $B(s) \neq \emptyset$ because any continuous function defined over a compact set has a maximum value (Weierstrass theorem). (iii) $B(s)$ is a convex set for each s because, by the quasiconcavity, $B_i(s_{-i})$ is a convex set for any s. (iv) The graph of B is a closed subset of $S \times S$. By the assumed continuity of u_i for each i, the verification holds by the proof used for Theorem 1.19. ■

Finally, the following is an extension of Nash's theorem for games with possibly infinitely many actions per player. The proof is essentially the same as the proof of Nash's theorem. Note that the conditions do not require quasi-concavity of the value functions as in the conditions for Theorem 1.22, but the conclusion is weaker than the conclusion of Theorem 1.22 because only the existence of a mixed strategy Nash equilibrium is guaranteed.

Theorem 1.23 *(Glicksberg) Consider a normal form game $(I, (S_i)_{i \in I}, (u_i)_{i \in I})$ such that to I is finite, the sets S_i are nonempty, compact metric spaces, and the payoff functions $u_i : S \to \mathbb{R}$ are continuous, where $S = S_1 \times \cdots \times S_n$. Then a mixed strategy Nash equilibrium exists.*

Example 1.24 *Consider the two-person zero sum game such that each player selects a point on a circle of circumference one. Player 1 wants to minimize the distance (length of shorter path along the circle) between the two points, and Player 2 wants to maximize it. No pure strategy Nash equilibrium exists. Theorem* (1.23) *ensures the existence of a mixed strategy equilibrium. There are many.*

1.5 On the uniqueness of a Nash equilibrium

Consider a normal form (aka strategic form) game $(I, (S_i), (u_i))$ with the following notation and assumptions:

- $I = \{1, \ldots, n\} = [n]$, indexes the players
- $S_i \subset \mathbb{R}^{m_i}$ is action space of player i, assumed to be a convex set
- $S = S_1 \times \cdots \times S_n$.
- $u_i : S \to \mathbb{R}$ is the payoff function of player i. Suppose $u_i(x_i, x_{-i})$ is differentiable in x_i for x_i in an open set containing S_i for each i and x_{-i},

Definition 1.25 *The set of payoff functions $u = (u_1, \ldots, u_n)$ is* diagonally strictly concave *(DSC) if for every $x^*, \overline{x} \in S$ with $x^* \neq \overline{x}$,*

$$\sum_{i=1}^{n} \left((\bar{x}_i - x_i^*) \cdot \nabla_{x_i} u_i(x^*) + (x_i^* - \bar{x}_i) \cdot \nabla_{x_i} u_i(\bar{x}) \right) > 0. \tag{1.6}$$

Lemma 1.26 *If u is DSC then $u_i(x_i, x_{-i})$ is a concave function of x_i for i and x_{-i} fixed.*

Proof. Suppose u is DSC. Fix $i \in I$ and $\bar{x}_{-i} \in S_{-i}$. Suppose $\bar{x}_i$ and x_i^* vary while $\bar{x}_{-i}$ and x_{-i}^* are both fixed and set to be equal: $\bar{x}_{-i} = x_{-i}^*$. Then the DSC condition yields

$$(\bar{x}_i - x_i^*) \cdot \nabla_{x_i} u_i(x_i^*, \bar{x}_{-i}) > (\bar{x}_i - x_i^*) \cdot \nabla_{x_i} u_i(\bar{x}_i, \bar{x}_{-i}),$$

which means the derivative of the function $u_i(\cdot, \bar{x}_{-i})$ along the line segment from x_i^* to $\bar{x}_i$ is strictly decreasing, implying the conclusion. ■

Remark 1.27 *If u_i for each i depends only on x_i and is strongly concave in x_i then the DSC condition holds. Other examples of DSC functions can be obtained by selecting functions u_i that are strongly concave with respect to x_i and weakly dependent on x_{-i}.*

Theorem 1.28 *(Sufficient condition for uniqueness of Nash equilibrium [Rosen [19]]) Suppose u is DSC. Then there exists at most one Nash equilibrium. If, in addition, the sets S_i are closed and bounded (i.e. compact) and the functions $u_1, \ldots, u_n$ are continuous, there exists a unique Nash equilibrium.*

Proof. Suppose u is diagonally strictly concave and fix $x^*, \bar{x} \in S$. If x^* is a Nash equilibrium point then by definition, for each $i \in I$, $x_i^* \in \arg\max_{x_i} u_i(x_i, x_{-i}^*)$. By Lemma 1.26, the function $x_i \mapsto u_i(x_i, \bar{x}_{-i})$ is concave. By Proposition 1.34 on the first order optimality conditions for the maximum of a convex function, $(\bar{x}_i - x_i^*) \cdot \nabla_{x_i} u_i(x^*) \leq 0$. Similarly, if $\bar{x}$ is a Nash equilibrium, $(x_i^* - \bar{x}_i) \cdot \nabla_{x_i} u_i(\bar{x}) \leq 0$. Thus, for each i, both terms in the sum on the lefthand side of (1.6) are less than or equal to zero, in contradiction of (1.6). Therefore, there can be at most one Nash equilibrium point.

If, in addition, the sets S_i are compact and the functions u_i are continuous, in view of Lemma 1.26, the sufficient conditions in Theorem 1.22 hold, so there exists a Nash equilibrium, which is unique as already shown. ■

There is a sufficient condition for u to be DSC that is expressed in terms of some second order derivatives. Let U denote the $m \times m$ matrix, where $m = m_1 + \cdots + m_n$:

$$U(x) = \begin{pmatrix} \frac{\partial^2 u_1}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 u_1}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 u_n}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 u_n}{\partial x_n \partial x_n} \end{pmatrix}.$$

The matrix $U(x)$ has a block structure, where the ij^{th} block is the $m_i \times m_j$ matrix $\frac{\partial^2 u_i}{\partial x_i \partial x_j}$.

Corollary 1.29 *If the derivatives in the definition of $U(x)$ exist and if $U(x)+U(x)^T \prec 0$ (i.e. $U(x)+U^T(x)$ is strictly negative definite) for all $x \in S$, then u is DSC. In particular, there can be at most one Nash equilibrium.*

Proof. Note that for each i,

$$\begin{aligned} \nabla_{x_i} u_i(x^*) - \nabla_{x_i} u_i(\bar{x}) &= \nabla_{x_i} u_i(\bar{x} + t(x^* - \bar{x}))\Big|_{t=0}^{1} \\ &= \int_0^1 \frac{d}{dt} \left\{ \nabla_{x_i} u_i(\bar{x} + t(x^* - \bar{x})) \right\} dt \\ &= \int_0^1 \sum_{j=1}^{n} \frac{\partial \left(\nabla_{x_i} u_i \right)}{\partial x_j} (\bar{x} + t(x^* - \bar{x}))(x_j^* - \bar{x}_j) dt \end{aligned}$$

Multiplying on the right by $(\bar{x}_i - x_i^*)^T$ and summing over i yields that the lefthand side of (1.6) is equal to

$$\int_0^1 (\bar{x}-x^*)^T U(\bar{x}+t(x^*-\bar{x}))(x^*-\bar{x})dt = -\frac{1}{2}\int_0^1 (\bar{x}-x^*)^T (U+U^T)\Big|_{\bar{x}+t(x^*-\bar{x})}(\bar{x}-x^*)dt, \tag{1.7}$$

where $\bar{x}$ and x^* are viewed as vectors in $\mathbb{R}^{m_1+\cdots+m_n}$. Thus, if $U(x)+U^T(x)$ is strictly negative definite for all x, then the integrand on the righthand side of (1.7) is strictly negative, implying u is DSC. ■

Remark 1.30 *The following fact is sometimes helpful in applying Corollary 1.29. A sufficient condition for a symmetric matrix to be negative definite is that the diagonal elements are strictly negative, and the sum of the absolute values of the off-diagonal elements in any row is strictly smaller than the absolute value of the diagonal element in the row.*

Example 1.31 *Consider the two-player identical interests game with* $S_1 = S_2 = \mathbb{R}$ *and* $u_1(s_1,s_2) = u_2(s_1,s_2) = -\frac{1}{2}(s_1-s_2)^2$. *The Nash equilibria consist of policy profiles of the form* (s_1,s_1) *such that* $s_1 \in \mathbb{R}$. *So the Nash equilibrium is not unique. The matrix* $U(x)$ *is given by* $U(x) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$. *While* $U(x)+U(x)^T$ *is negative semidefinite, it is not (strictly) negative semidefinite, which is why Corollary 1.29 does not hold.*

If instead $u_1(s_1,s_2) = -\frac{a}{2}s_1^2$ *and* $u_2(s_1,s_2) = -\frac{1}{2}(s_1-s_2)^2$ *for some* $a>0$, *then* $(0,0)$ *is a Nash equilibrium. Since* $U(x)+U(x)^T = \begin{pmatrix} -a & 0 \\ 1 & -1 \end{pmatrix} + \begin{pmatrix} -a & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -2a & 1 \\ 1 & -2 \end{pmatrix}$, *which is negative definite*[2] *if* $a > \frac{1}{4}$, *the Nash equilibrium is unique if* $a > \frac{1}{4}$, *by Corollary 1.29. Uniqueness can be easily checked directly for any* $a > 0$.

1.6 Two-player zero sum games

1.6.1 Saddle points and the value of two-player zero sum game

A two-player game is a zero sum game if the sum of the payoff functions is zero for any pair of actions taken by the two players. For such a game we let $\ell(p,q)$ denote the payoff function of player 2 as a function of the action p of player 1 and the action q of player 2. Thus, the payoff function of player 1 is $-\ell(p,q)$. In other words, the objective of player 1 is to minimize $\ell(p,q)$. So we can think of $\ell(p,q)$ as a loss function for player 1 and a reward function for player 2. A Nash equilibrium in this context is called a *saddle point.* In other words, by definition, $(\bar{p},\bar{q})$ is a saddle point if

$$\inf_p \ell(p,\bar{q}) = \ell(\bar{p},\bar{q}) = \sup_q \ell(\bar{p},q). \tag{1.8}$$

We say that $\bar{p}$ is minmax optimal for ℓ if $\sup_q \ell(\bar{p},q) = \inf_p \sup_q \ell(p,q)$, and similarly $\bar{q}$ is maxmin optimal for ℓ if $\inf_p \ell(p,\bar{q}) = \sup_q \inf_p \ell(p,q)$, We list a series of facts.

Fact 1 If p and q each range over a compact convex set and ℓ is jointly continuous in (p,q), quasiconvex in p and quasiconcave in q then there exits a saddle point. This follows from the Debreu, Glicksberg, Fan theorem based on fixed points, Theorem 1.22.

[2] A symmetric matrix A is negative definite if and only if $-A$ is positive definite. A 2×2 symmetric matrix is positive definite if and only if its diagonal elements and its determinant are positive.

Fact 2 For any choice of the function ℓ, weak duality holds:

$$\sup_q \inf_p \ell(p,q) \leq \inf_p \sup_q \ell(p,q). \tag{1.9}$$

Possible values of the righthand or lefthand sides of (1.9) are ∞ or $-\infty$. To understand (1.9), suppose player 1 is trying to select p to minimize ℓ and player 2 is trying to select q to maximize ℓ. The lefthand side represents the result if for any choice q of player 2, player 1 can select p depending on q. In other words, since "$\inf_p$" is closer to the objective function than "$\sup_q$," the player executing "$\inf_p$" has an advantage over the other player. The righthand side has the order of optimizations reversed, so the player executing the "$\sup_q$" operation has the advantage of knowing p. The number $\triangle \triangleq \inf_p \sup_q \ell(p,q) - \sup_q \inf_p \ell(p,q)$ is known as the *duality gap.* If both sides of (1.9) are ∞ or if both sides are $-\infty$ we set $\triangle = 0$. Hence, for any ℓ, the duality gap is nonnegative.

Here is a proof of (1.9). It suffices to show that for any finite constant c such that $\inf_p \sup_q \ell(p,q) < c$, it also holds that $\sup_q \inf_p \ell(p,q) < c$. So suppose c is a finite constant such that $\inf_p \sup_q \ell(p,q) < c$. Then by the definition of infimum there exists a choice $\bar{p}$ of p such that $\sup_q \ell(\bar{p},q) < c$. Clearly $\inf_p \ell(p,q) \leq \ell(\bar{p},q)$ for any q, so taking a supremum over q yields $\sup_q \inf_p \ell(p,q) \leq \sup_q \ell(\bar{p},q) < c$, as needed.

Fact 3 A pair $(\bar{p},\bar{q})$ is a saddle point if and only if $\bar{p}$ is minmax optimal, $\bar{q}$ is maxmin optimal, and there is no duality gap. Here is a proof.

(if) Suppose $\bar{p}$ is minmax optimal, there is no duality gap, and $\bar{q}$ is maxmin optimal. Using these properties in the order listed yields

$$\sup_q \ell(\bar{p},q) = \inf_p \sup_q \ell(p,q) = \sup_q \inf_p \ell(p,q) = \inf_p \ell(p,\bar{q}). \tag{1.10}$$

Since $\sup_q \ell(\bar{p},q) \geq \ell(\bar{p},\bar{q}) \geq \inf_p \ell(p,\bar{q})$, (1.10) implies $\sup_q \ell(\bar{p},q) = \ell(\bar{p},\bar{q}) = \inf_p \ell(p,\bar{q})$. In other words, $(\bar{p},\bar{q})$ is a saddle point.

(only if) Suppose $(\bar{p},\bar{q})$ is a saddle point, which by definition, means (1.8) holds. Extending (1.8) yields:

$$\sup_q \inf_p \ell(p,q) \geq \inf_p \ell(p,\bar{q}) = \ell(\bar{p},\bar{q}) = \sup_q \ell(\bar{p},q) \geq \inf_p \sup_q \ell(p,q). \tag{1.11}$$

By weak duality the far lefthand side of (1.11) is less than or equal to the far righthand side of (1.11), so the inequalities in (1.11) hold with equality. Thus, there is no duality gap, $\bar{p}$ is minmax optimal, and $\bar{q}$ is maxmin optimal.

Fact 4 For bilinear two-player zero-sum game with ℓ of the form $\ell(p,q) = pAq^T$, p and q are stochastic row vectors, the minmax problem for player 1 and maximn problem for player 2 are equivalent to dual linear programing problems.

A min-max strategy for player 1 is to select p to solve the problem:

$$\min_{p\in\triangle} \max_{q\in\triangle} pAq^T$$

We can formulate this minimax problem as a linear programing problem, which we view as the primal problem:

$$\min_{p,t:pA\leq t\mathbf{1}\ p\geq 0\ p\mathbf{1}=1} t \quad \text{(primal problem)}$$

Linear programming problems have no duality gap. To derive the dual linear program we consider the Lagrangian and switch the order of optimizations:[3]

$$\begin{aligned}
\min_{p\in\triangle}\max_{q\in\triangle} pAq^T &= \min_{p,t:pA\leq t\mathbf{1}\ p\geq 0\ p\mathbf{1}=1} t \\
&= \min_{p,t:\ p\geq 0}\ \max_{\lambda,\mu:\lambda\geq 0}\ t + (pA - t\mathbf{1}^T)\lambda + \mu(1-p\mathbf{1}) \\
&\qquad\quad \searrow \quad \swarrow \\
&= \max_{\lambda,\mu:\lambda\geq 0}\ \min_{p,t:\ p\geq 0}\ t + (pA - t\mathbf{1}^T)\lambda + \mu(1-p\mathbf{1}) \\
&= \max_{\mu,\lambda:\lambda\geq 0,\ A\lambda\geq\mu\mathbf{1},\lambda\mathbf{1}=1} \mu
\end{aligned}$$

In other words, the dual linear programming problem is

$$\max_{\mu,\lambda:\lambda\geq 0,\ A\lambda\geq\mu\mathbf{1},\lambda\mathbf{1}=1} \mu \qquad \text{(dual problem)}$$

which is equivalent to $\max_{\lambda\in\triangle}\min_{p\in\triangle} pA\lambda^T$, which is the maxmin problem for player 2. Thus, both the minmax problem of player 1 and the maxmin problem of player 2 can be formulated as linear programming problems, and those are dual problems. (See Section 1.7 below on KKT conditions and duality.)

1.7 Appendix: Derivatives, extreme values, and convex optimization

1.7.1 Weierstrass extreme value theorem

Suppose f is a function mapping a set S to $\mathbb{R}$. A point $x^*\in S$ is a maximizer of f if $f(x)\leq f(x^*)$ for all $x\in S$. The set of all maximizers of f over S is denoted by $\arg\max_{x\in S} f(x)$. It holds that $\arg\max_{x\in S} f(x) = \{x\in S : f(x) = \sup_{y\in S} f(y)\}$. It is possible that there are no maximizers.

Theorem 1.32 *(Weierstrass extreme value theorem) Suppose $f : S\to\mathbb{R}$ is a continuous function and the domain S is a sequentially compact set. (For example, S could be a closed, bounded subset of $\mathbb{R}^m$ for some m.) Then there exists a maximizer of f. In other wordsd,* $\arg\max_{x\in S} f(x)\neq\emptyset$.

Proof. Let $V = \sup_{x\in S} f(x)$. Note that $V\leq\infty$. Let (x_n) denote a sequence of points in S such that $\lim_{n\to\infty} f(x_n) = V$. By the compactness of S, there is a subsequence (x_{n_k}) of the points that is convergent to some point $x^*\in S$. In other words, $\lim_{k\to\infty} x_{n_k} = x^*$. By the continuity of f, $f(x^*) = \lim_{k\to\infty} f(x_{n_k})$, and also the subsequence of values has the same limit as the entire sequence of values, so $\lim_{k\to\infty} f(x_{n_k}) = V$. Thus, $f(x^*) = V$, which implies the conclusion of the theorem. ∎

[3] For readability, we consider the condition $p\geq 0$ on p to define the domain X for the primal problem, instead of introducing a Lagrange multiplier for it. However, for this problem, we would get the same dual if we instead introduced another multiplier λ' for the constraint $p\geq 0$, yielding an additional term $-p\lambda'$ in the Lagrangian, where λ' is a column vector with $\lambda'\geq 0$, and the min is over all p, not just $p\geq 0$. After eliminating p from the minimization problem, the dual problem is to maximize μ over $\mu,\lambda\geq 0,\lambda'\geq 0$ subject to $A\lambda-\mu 1=\lambda'$ and $\lambda\mathbf{1}=1$. Then we can easily remove the variable λ' to get the form shown.

Example 1.33 *(a) If $S = [0, 1)$ and $f(x) = x^2$ there is no maximizer. Theorem 1.32 doesn't apply because S is not compact.*

(b) If $S = \mathbb{R}$ and $f(x) = x^2$ there is no maximizer. Theorem 1.32 doesn't apply because S is not compact.

(c) If $S = [0, 1]$ and $f(x) = x$ for $0 \le x < 0.5$ and $f(x) = 0$ for $0.5 \le x \le 1$ then there is no maximizer. Theorem 1.32 doesn't apply because f is not continuous.

1.7.2 Derivatives of functions of several variables

Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$. We say that f is differentiable at a point x if f is well enough approximated in a neighborhood of x by a linear approximation. Specifically, an $m \times n$ matrix $J(x)$ is the Jacobian of f at x if

$$\lim_{a \to x} \frac{\|f(a) - f(x) - J(x)(a - x)\|}{\|a - x\|} = 0$$

The Jacobian is also denoted by $\frac{\partial f}{\partial x}$ and if f is differentiable at x the Jacobian is given by a matrix of partial derivatives:

$$\frac{\partial f}{\partial x} = J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Moreover, according to the multidimensional differentiability theorem, a sufficient condition for f to be differentiable at x is for the partial derivatives $\frac{\partial f_i}{\partial x_j}$ to exist and be continuous in a neighborhood of x. In the special case $m = 1$ the gradient is the transpose of the derivative:

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at x if there is an $n \times n$ matrix $H(x)$, called the Hessian matrix, such that

$$\lim_{a \to x} \frac{\|f(a) - f(x) - J(x) \cdot (a - x) - \frac{1}{2}(a - x)^T H(x)(a - x)\|}{\|a - x\|^2} = 0.$$

The matrix $H(x)$ is also denoted by $\frac{\partial^2 f}{(\partial x)^2}(x)$, and is given by a matrix of second order partial derivatives:

$$\frac{\partial^2 f}{(\partial x)^2} = H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

The function f is twice differentiable at x if both the first partial derivatives $\frac{\partial f}{\partial x_i}$ and second order partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exist and are continuous in a neighborhood of x.

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and if $x, \alpha \in \mathbb{R}^n$, then we can find the first and second derivatives of the function $t \mapsto f(x + \alpha t)$ from $\mathbb{R} \to \mathbb{R}$:

$$\begin{aligned}
\frac{\partial f(x+\alpha t)}{\partial t} &= \sum_i \left.\frac{\partial f}{\partial x_i}\right|_{x+\alpha t} \alpha_i = \alpha^T \nabla f(x+\alpha t). \\
\frac{\partial^2 f(x+\alpha t)}{(\partial t)^2} &= \sum_i \sum_j \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{x+\alpha t} \alpha_i \alpha_j \\
&= \alpha^T H(x+\alpha t)\alpha.
\end{aligned}$$

If $H(y)$ is positive semidefinite for all y, in other words $\alpha^T H(y)\alpha \geq 0$ for all $\alpha \in \mathbb{R}^n$ and all y, then f is a convex function.

1.7.3 Optimality conditions for convex optimization

(See, for example, [1], for a more extensive version of this section.) A subset $C \subset \mathbb{R}^n$ is convex if whenever $x, x' \in C$ and $0 \leq \lambda \leq 1$, $\lambda x + (1-\lambda)x' \in C$. The affine span of a set $C \subset \mathbb{R}^n$ is defined by $\text{aff}(C) = \{\sum_{i=1}^k \lambda_i x_i : k \geq 1, x_i \in C, \sum_i \lambda_i = 1\}$. The relative interior of a convex set C is the set of $x \in C$ such that for some $\epsilon > 0$, $B(\epsilon, x) \cap \text{aff}(C) \subset C$, where $B(\epsilon, x)$ is the radius ϵ ball centered at x.

A function $f : C \to \mathbb{R}$ is convex if whenever $x, x' \in C$ and $0 \leq \lambda \leq 1$, $f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda) f(x')$. If f is differentiable over an open convex set C, then f is convex if and only for any $x, y \in C$, $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$. If f is twice differentiable over an open convex set C, it is convex if and only if the Hessian is positive semidefinite over C, i.e. $H(x) \succeq 0$ for $x \in C$.

Proposition 1.34 *(First order optimality condition for convex optimization) Suppose f is a convex differentiable function on a convex open domain D so that $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ for all $x, y \in D$. Suppose C is a convex set with $C \subset D$. Then $x^* \in \arg\min_{x \in C} f(x)$ if and only if $(y - x^*) \cdot \nabla f(x^*) \geq 0$ for all $y \in C$.*

Proof. (if) If $(y - x^*) \cdot \nabla f(x^*) \geq 0$ for all $y \in C$, then for any $y \in C$, $f(y) \geq f(x^*) + \nabla f(x^*) \cdot (y - x^*) \geq f(x^*)$, so $x^* \in \arg\min_{x \in C} f(x)$.

(only if) Conversely, suppose $x^* \in \arg\min_{x \in C} f(x)$ and let $y \in C$. Then for any $\lambda \in (0, 1)$, $(1-\lambda)x^* + \lambda y \in C$ so that $f(x^*) \leq f((1-\lambda)x^* + \lambda y) = f(x^* + \lambda(y - x^*))$. Thus, $\frac{f(x^* + \lambda(y - x^*)) - f(x^*)}{\lambda} \geq 0$ for all $\lambda > 0$. Taking $\lambda \to 0$ yields $(y - x^*) \cdot \nabla f(x^*) \geq 0$. ■

Next we discuss the Karush-Kuhn-Tucker necessary conditions for convex optimization, involving multipliers for constraints. Consider the optimization problem

$$\begin{aligned}
\min_{x \in X} \;\; & f(x) \\
\text{s.t.} \;\; & g_i(x) \leq 0; \;\; i \in [m] \\
& h_j(x) = 0; \;\; j \in [\ell],
\end{aligned} \tag{1.12}$$

such that $X \subset \mathbb{R}^n$, $f : X \to \mathbb{R}$ is the objective function, inequality constraints are expressed in terms of the functions $g_i : X \to \mathbb{R}$ and equality constraints are expressed in terms of the functions $h_j : X \to \mathbb{R}$.

The optimization problem is *convex* if X is a convex set, the function f and the g_i's are convex and the h_j's are affine. The optimization problem satisfies the (relaxed) *Slater condition* if it is a convex optimization problem and there exists an $\bar{x}$ that is strictly feasible: $\bar{x}$ is in the relative interior of X, $g_i(\bar{x}) < 0$ for all i

such that g_i is not an affine function, $g_i(\bar{x}) \le 0$ for all i such that g_i is an affine function, and $h_j(\bar{x}) = 0$ for all j.

Given real valued multipliers $\lambda_i, i \in [m]$, and $\mu_j, j \in [\ell]$, the Lagrangian function is defined by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x).$$

which we also write as: $L(x, \lambda, \mu) = f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle$.

Theorem 1.35 *(Karush-Kuhn-Tucker (KKT) necessary conditions) Consider the optimization problem* (1.12) *such that f, the g_i's and the h_j's are continuously differentiable in a neighborhood of a point x^*. If x^* is a local minimum and a regularity condition is satisfied (e.g., linearity of constraint functions, or linear independence of the gradients of the active inequality constraints and the equality constraints, or the (relaxed) Slater condition holds) then there exist $\lambda_i, i \in [m]$, and $\mu_j, j \in [\ell]$, called Lagrange multipliers, such that the following conditions, called the KKT conditions, hold:*

(gradient of Lagrangian with respect to x is zero)

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{\ell} \mu_j \nabla h_j(x^*) = 0 \tag{1.13}$$

(primal feasibility)

$$x^* \in X \tag{1.14}$$
$$g_i(x^*) \le 0;\ i \in [m] \tag{1.15}$$
$$h_j(x^*) = 0;\ j \in [\ell] \tag{1.16}$$

(dual feasibility)

$$\lambda_i \ge 0;\ \ i \in [m] \tag{1.17}$$

(complementary slackness)

$$\lambda_i g_i(x^*) = 0;\ \ i \in [m] \tag{1.18}$$

Theorem 1.36 *(Karush-Kuhn-Tucker sufficient conditions) Suppose the optimization problem* (1.12) *is convex and the g_i's are continuously differentiable (convex) functions. If x^*, $(\lambda_i)_{i\in[m]}$, and $(\mu_j)_{j\in[\ell]}$, satisfy the KKT conditions* (1.13)-(1.18)*, then x^* is a solution of* (1.12)*.*

Corollary 1.37 *If the primal problem is linear, or if it is convex, the g_i's are differentiable, and the Slater condition is satisfied, then x^* is a solution of* (1.12) *if and only if there exits (λ, μ) such that the KKT conditions* (1.13)-(1.18) *are satisfied.*

The following describes the dual of the above problem in the convex case. Suppose that the optimization problem (1.12) is convex. The dual objective function ϕ is defined by

$$\begin{aligned}\phi(\lambda, \mu) &= \min_{x\in X} L(x, \lambda, \mu)\\ &= \min_{x\in X} f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle,\end{aligned}$$

The dual optimization problem can be expressed as

$$\begin{aligned} &\max_{\lambda,\mu} \phi(\lambda,\mu) \\ &\quad \text{s.t.} \quad \lambda_i \geq 0; i \in [m] \end{aligned} \tag{1.19}$$

In general, the optimal value of the dual optimization problem is less than or equal to the optimal value of the primal problem (this property is called *weak duality*), because:

$$\begin{aligned} \min_{x\in X: g(x)\leq 0, h(x)=0} f(x) &= \min_{x\in X} \max_{\lambda,\mu:\lambda\geq 0} L(x,\lambda,\mu) \\ &\geq \max_{\lambda,\mu:\lambda\geq 0} \min_{x\in X} L(x,\lambda,\mu) \\ &= \max_{\lambda,\mu:\lambda\geq 0} \phi(\lambda,\mu). \end{aligned}$$

The duality gap is the optimal primal value minus the optimal dual value. Thus, the duality gap is always greater than or equal to zero. *Strong duality* is said to hold if the duality gap is zero. The duality gap is zero if the primal, and hence the dual, problems are linear.

The duality gap is also zero if the primal problem is convex, the g_i's are differentiable, and there exists x^*, λ, μ satisfying the KKT conditions. To see why, note that $L(x, \lambda, \mu)$ is convex in x by dual feasibility and convexity of the g_i's, and its gradient with respect to x is zero at x^*. So

$$\phi(\lambda,\mu) \triangleq \min_{x\in X} L(x,\lambda,\mu) = L(x^*,\lambda,\mu) = f(x^*) + \langle \lambda, g(x^*)\rangle + \langle \mu, h(x^*)\rangle = f(x^*),$$

where $\langle \lambda, g(x^*)\rangle = 0$ by complementary slackness and $\langle \mu, h(x^*)\rangle = 0$ by primal feasibility. Finally, note that the existence of primal feasible x^* and dual feasible λ, μ such that $\phi(\lambda,\mu) = f(x^*)$ implies strong duality.

Example 1.38 *Suppose a factory has an inventory with various amounts of commodities (raw materials). Specifically, it has C_i units of commodity i for each i. The factory is capable of producing several different goods, with market price p_j per unit of good j. Suppose producing one unit of good j requires A_{ij} units of commodity i for each i. How could the factory maximize the value of its inventory? It could decide to produce x_j units of good j, where the x's are selected to maximize the total selling price of the goods, subject to the constraint on needed resources. Given C, p, and A, this can be formulated as a linear programming problem:*

$$\begin{aligned} &\max \ p^T x \\ &\quad \text{over } x \geq 0 \\ &\quad \text{subject to } Ax \leq C. \end{aligned}$$

We derive the dual problem by introducing a multiplier vector λ for the constraint $Ax \leq C$. We shall consider the condition $x \geq 0$ to define the domain X of the problem, instead of introducing a multiplier for it in the Lagrangian. The Lagrangian is $p^T x + \lambda^T(C - Ax)$ and the dual cost function is $\max_{x\geq 0} \lambda^T C + (p^T - \lambda^T A)x = \lambda^T C$, as long as $\lambda^T A \geq p^T$; otherwise the dual cost is infinite. Thus, the dual problem is

$$\begin{aligned} &\min \ \lambda^T C \\ &\quad s.t \quad \lambda \geq 0 \\ &\quad\quad \lambda^T A \geq p^T. \end{aligned}$$

The dual problem offers a second way to compute the same value for the inventory. Think of λ_i as a value per unit of commodity i. An interpretation of the constraint $\lambda^T A \geq p^T$ is that the sum of the values of the *commodities used for any good should be at least as large as the price of that good, on a per unit basis. So a potential buyer of the inventory could argue that a vector of commodity prices λ would be a fair price to pay to the factory for the inventory, because for any good, the sum of the prices of the commodities needed to produce one unit of good is greater than or equal to the unit price of the good. In other words, if one unit of good of any type j were purchased at the market price p_j, and the good could be decomposed into its constituent commodities, then the value of those commodities for price vector λ would be greater than or equal p_j.*

Example 1.39 *Here is a classic example violating Slater's condition for $x \in \mathbb{R}^2$. Let $f(x) = x_2$, $g_1(x) = (x_1-1)^2 + x_2^2 - 1$ and $g_2(x) = (x_1+1)^2 + x_2^2 - 1$. Consider the problem of minimizing $f(x)$ subject to $g_1(x) \le 0$ and $g_2(x) \le 0$. Geometrically speaking, the feasible set is the intersection of two radius one disks, with centers at $\binom{\pm 1}{0}$. The disks intersect at the single point $x^* = \binom{0}{0}$, which is hence the only feasible point, and which is thus also the minimizer. Note that $\nabla L\Big|_{x^*} = \binom{0}{1} + \lambda_1\binom{-2}{0} + \lambda_2\binom{2}{0}$, so there is no choice of λ_1 and λ_2 to make $\nabla L(x^*, \lambda_1, \lambda_2) = 0$.*

Example 1.40 *Consider the problem $\min_{x\le 0} f(x)$ for some extended real-valued function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$. Letting $g(x) = x$, we see the Lagrangian is given by $L(x, \lambda) = f(x) + \lambda x$, and the dual function is given by $\phi(\lambda) = \min_{x\le 0} f(x) + \lambda x$. Note that for any fixed x and λ, the value of $f(x) + \lambda x$ is the y-intercept of the line through $(x, f(x))$ in the (x, y) plane, with slope $-\lambda$. Thus, for each $\lambda \ge 0$, $\phi(\lambda)$ is the minimum y-intercept over all lines with slope $-\lambda$ that intersect the graph of f. Examples based on two different choices of f are shown in Figure 1.4. For each one, a line of slope $-\lambda$ is shown for some $\lambda > 0$, such that the y intercept is as small as possible for the given λ. In other words, the y intercept is $\phi(\lambda)$. For case (a), ϕ is maximized over $\lambda \ge 0$ at $\lambda = -f'(0)$. For case (b), $\phi(\lambda)$ is maximized over $\lambda \ge 0$ at $\lambda = 0$. Since $g(x) < 0$ for $x < 0$, Slater's condition is satisfied for both cases.*

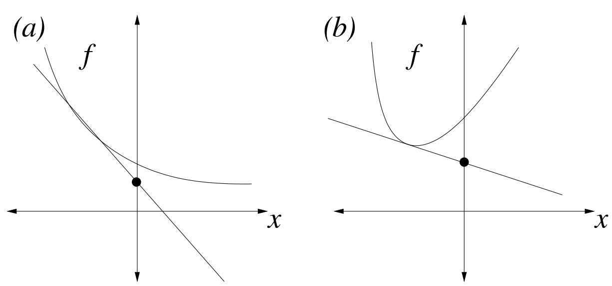


Figure 1.4: Illustration of $\phi(\lambda)$.

Chapter 2

Evolution as a Game

2.1 Evolutionarily stable strategies

Consider a population of individuals, where each individual is of some type. Suppose individuals have occasional pairwise encounters. During an encounter the two individuals involved play a two-player symmetric game in which the strategies are the types of the individuals. As a result of the encounter, each of the two individuals produces a number of offspring of its same type, with the number being determined by a fitness table or, equivalently, a fitness matrix. For example, consider a population of crickets such that each cricket is either small or large. If two small crickets meet each other then they each spawn five more small crickets. If a small cricket encounters a large cricket then the small cricket spawns one more small cricket and the large cricket spawns eight new large crickets. If two large crickets meet then each of them spawns three new large crickets. We can summarize these outcomes using the fitness matrix shown in Table 2.1.

Table 2.1: Fitness matrix for a population consisting of small and large crickets.

	small	*large*
small	5, 5	1, 8
large	8, 1	3, 3

or, for short, $u = \begin{pmatrix} 5 & 1 \\ 8 & 3 \end{pmatrix}$.

If a type i individual encounters a type j individual, then the type i individual spawns $u(i,j)$ new individuals of type i, and the type j individual spawns $u(j,i)$ new individuals of type j.

For example, consider a homogeneous population in which all individuals are of a single type S. Suppose a small number of individuals of type T is introduced into the population (or some individuals of type T invade the population). If the T's were to replicate faster than the S's they could change the composition of the population. For example, that is why conservationists are seeking to prevent an invasive species of fish from entering Lake Michigan `http://www.chicagotribune.com/news/nationworld/midwest/ct-asian-carp-lake-michigan-20170623-story.html`. Roughly speaking, the type S is said to be evolutionarily stable if S is not susceptible to such invasions, in the regime of very large populations.

Definition 2.1 *(Evolutionarily stable pure strategies) A type S is an evolutionarily stable strategy (ESS) if* *for all $\epsilon > 0$ sufficiently small, and any other type T,*

$$(1-\epsilon)u(S,S) + \epsilon u(S,T) > (1-\epsilon)u(T,S) + \epsilon u(T,T)$$

In other words, if S is invaded by T at level ϵ then S has a strictly higher mean fitness level than T.

Note that the ESS property is determined entirely based on the fitness matrix; no explicit population dynamics are involved in the definition.

Consider the large vs. small crickets example with fitness matrix given by Table 2.1. The large type is ESS by the following observations. For a population of large crickets with a level ϵ invasion of small crickets, the average fitness of a large cricket is $3(1-\epsilon) + 8\epsilon = 3 + 5\epsilon$, while the average fitness of a small cricket is $(1-\epsilon) + 5\epsilon = 1 + 4\epsilon$. So the invading small crickets are less fit, suggesting their population will stay small compared to the population of large crickets. In contrast, small is not ESS for this example. For a population of small crickets with a level ϵ invasion of large crickets, the average fitness of a small cricket is $5(1-\epsilon) + \epsilon = 5 - 4\epsilon$, while the average fitness of a large cricket is $8(1-\epsilon) + 3\epsilon = 8 - 5\epsilon$. Thus, the average fitness of the invading large crickets is greater than the average fitness of the small crickets. Note that for the bi-matrix game specified in Table 2.1, large is a strictly dominant strategy (or type). In general, strictly dominant strategies are ESS and if there is a strictly dominant strategy, no other strategy can be ESS.

The definition of ESS can be extended to mixed strategies, as follows.

Definition 2.2 *(Evolutionarily stable mixed strategies) A mixed strategy p^* is an evolutionarily stable strategy (ESS) if there is an $\bar{\epsilon} > 0$ so that for any ϵ with $0 < \epsilon \leq \bar{\epsilon}$ and any mixed strategy p' with $p' \neq p^*$,*

$$u(p^*, (1-\epsilon)p^* + \epsilon p') > u(p', (1-\epsilon)p^* + \epsilon p'). \tag{2.1}$$

Proposition 2.3 *(First characterization of ESS using Maynard Smith condition) p^* is ESS if and only if there exists $\widehat{\epsilon} > 0$ such that*

$$u(p^*, \bar{p}) > u(\bar{p}, \bar{p}) \tag{2.2}$$

for all $\bar{p}$ with $0 < \|p^* - \bar{p}\|_1 \leq 2\widehat{\epsilon}$. (By definition, $\|p^* - \bar{p}\|_1 = \sum_a |p^*_a - \bar{p}_a|$.)

Proof. Before proving the if and only if portions separately, note the following. Given strategies p^* and p' and $\epsilon > 0$, let $\bar{p} = (1-\epsilon)p^* + \epsilon p'$. Then $u(\bar{p}, \bar{p}) = u((1-\epsilon)p^* + \epsilon p', \bar{p}) = (1-\epsilon)u(p^*, \bar{p}) + \epsilon u(p', \bar{p})$, which implies that (2.1) and (2.2) are equivalent.

(if) (We can take $\widehat{\epsilon}$ and $\bar{\epsilon}$ to be the same for this direction.) Suppose there exists $\bar{\epsilon} > 0$ so that (2.2) holds for all $\bar{p}$ with $0 < \|p^* - \bar{p}\|_1 \leq 2\bar{\epsilon}$. Let p' be any strategy with $p' \neq p^*$ and let ϵ satisfy $0 < \epsilon \leq \bar{\epsilon}$. Let $\bar{p} = (1-\epsilon)p^* + \epsilon p'$. Then $0 < \|p^* - \bar{p}\|_1 \leq 2\bar{\epsilon}$ so that (2.2) holds, which is equivalent to (2.1), so that p^* is ESS.

(only if) Suppose p^* is ESS. Let $\bar{\epsilon}$ be as in the definition of ESS so that (2.1) holds for any $p' \neq p^*$ and any ϵ with $0 < \epsilon \leq \bar{\epsilon}$. Let $\widehat{\epsilon} = \bar{\epsilon}\min\{p^*_i : p^*_i > 0\}$. Let $\bar{p}$ be a mixed strategy such that $0 < \|p^* - \bar{p}\|_1 \leq 2\widehat{\epsilon}$. In particular, $|p^*_i - \bar{p}_i| \leq \widehat{\epsilon}$ for all i. Then there exists a mixed strategy p' such that $\bar{p} = (1-\bar{\epsilon})p^* + \bar{\epsilon}p'$ for some p' with $p' \neq p^*$. Indeed, it must be that $p' = \frac{\bar{p} - (1-\bar{\epsilon})p^*}{\bar{\epsilon}}$. It is easy to check that the entries of p' sum to one. Furthermore, clearly $p'_i \geq 0$ if $p^*_i = 0$. If $p^*_i > 0$ then $\bar{p}_i - (1-\bar{\epsilon})p^*_i \geq p^*_i - \widehat{\epsilon} - (1-\bar{\epsilon})p^*_i = \bar{\epsilon}p^*_i - \widehat{\epsilon} \geq 0$. By assumption, (2.1) holds, which is equivalent to (2.2). ■

Proposition 2.4 *(Second characterization of ESS using Maynard Smith condition) p^* is ESS if and only if for every $p' \neq p^*$, either*

(i) $u(p^*, p^*) > u(p', p^*)$, or

(ii) $u(p^*, p^*) = u(p', p^*)$ and the Maynard Smith condition holds: $u(p^*, p') > u(p', p')$.

Proof. (only if) Suppose p^* is an ESS. Since $u(p, q)$ is linear in each argument, (2.1) is equivalent to

$$(1-\epsilon)u(p^*, p^*) + \epsilon u(p^*, p') > (1-\epsilon)u(p', p^*) + \epsilon u(p', p') \tag{2.3}$$

so there exists $\bar{\epsilon} > 0$ so that (2.3) holds for all $0 < \epsilon \leq \bar{\epsilon}$. Since the terms with factors $(1-\epsilon)$ dominate as $\epsilon \to 0$, it follows that either (i) or (ii) holds.

(if) (The proof for this part is slightly complicated because in the definition of ESS, the choice of $\bar{\epsilon}$ is independent of p'.) Suppose either (i) or (ii) holds for every $p' \neq p^*$. Then $u(p^*, p^*) \geq u(p', p^*)$ for all p'. Let $F = \{p' : u(p^*, p^*) = u(p', p^*)\}$ and let $G = \{p' : u(p^*, p') > u(p', p')\}$. By the continuity of u, F is a closed set and G is an open set, within the set of mixed strategies Σ. By assumption, $F \subset G$. The function $u(p^*, p^*) - u(p', p^*)$ is strictly positive on F^c and hence also on the compact set $G^c = \Sigma \backslash G$. Since G^c is a compact set, the minimum of $u(p^*, p^*) - u(p', p^*)$ over G^c exists, and is strictly positive. The function $p' \mapsto u(p^*, p') - u(p', p')$ is a continuous function on the compact set Σ and is thus bounded below by some possibly negative constant. So there exists $\bar{\epsilon} > 0$ such that

$$(1-\bar{\epsilon}) \min\{u(p^*, p^*) - u(p', p^*) : p' \in G^c\} + \bar{\epsilon} \min\{u(p^*, p') - u(p', p') : p' \in \Sigma\} > 0.$$

It follows that for any ϵ with $0 < \epsilon \leq \bar{\epsilon}$, (2.3), and hence also (2.1), holds. Thus, p^* is ESS. ■

The following is immediate from Proposition 2.4.

Corollary 2.5 *(ESS and Nash equilibria) Consider a symmetric two-player normal form game.*

(i) If a mixed strategy p is ESS then (p, p) is a Nash equilibrium.

(ii) If (p, p) is a strict Nash equilibrium in mixed strategies, then p is ESS.

2.2 Replicator dynamics

Continue to consider a symmetric two-player game with payoff functions u_1 and u_2. The symmetry means $S_1 = S_2$ and for any strategy profile (x, y), $u_1(x, y) = u_2(y, x)$, and $u_1(x, y)$ is the same as $u(x, y)$, where u is the fitness matrix. For brevity, we write $u(x, y)$ instead of $u_1(x, y)$. Consider a large population such that each individual in the population has a type in S_1. Let $\eta_t(a)$ denote the number of type a individuals at time t. We take $\eta_t(a)$ to be a nonnegative real value, rather than an integer. Assuming it is a large real value the difference is relatively small. Sometimes such models are called fluid models. Let $\theta_t(a) = \frac{\eta_t(a)}{\sum_{a'} \eta_t(a')}$, so that $\theta_t(a)$ is the fraction of individuals of type a. In other words, if an individual were selected from the population uniformly at random at time t, θ_t represents the probability distribution of the type of the individual. Recall that, thinking of u as the payoff of player 1 in a normal form game, $u(a, \theta_t)$ is the expected payoff of player 1 if player 2 uses the mixed strategy θ_t. In the context of evolutionary games, $u(a, \theta_t)$ is the average fitness of an individual of type a for an encounter with another individual selected uniformly at random from the population. The (continuous time, deterministic) replicator dynamics is given by the following ordinary differential equation, known as the *fitness equation*:

$$\dot{\eta}_t(a) = \eta_t(a)u(a, \theta_t).$$

The fitness equation implies an equation for the fractions. Let $D_t = \sum_{a'} \eta_t(a')$ so that $\theta_t(a) = \frac{\eta_t(a)}{D_t}$. By the fitness equation and the rule for derivatives of ratios,

$$\begin{aligned}\dot{\theta}_t(a) &= \frac{\dot{\eta}_t(a)D_t - \eta_t(a)\dot{D}_t}{D_t^2} \\ &= \frac{\eta_t(a)u(a,\theta_t)}{D_t} - \frac{\eta_t(a)\sum_{a'}\eta_t(a')u(a',\theta_t)}{D_t^2}\end{aligned}$$

which can be written as:

$$\dot{\theta}_t(a) = \theta_t(a)(u(a, \theta_t) - u(\theta_t, \theta_t)). \tag{2.4}$$

The term $u(\theta_t, \theta_t)$ in (2.4) is the average over the population of the average fitness of the population. Thus, the fraction of type a individuals increases if the fitness of that type against the population, namely $u(a, \theta_t)$, is greater than the average fitness of all types.

Let $\bar{\theta}$ be a population share state vector for the replicator dynamics. In other words, $\bar{\theta}$ is a probability vector over the finite set of types, S_1. The following definition is standard in the theory of dynamical systems:

Definition 2.6 *(Classification of states for the replicator dynamics)*

(i) A vector $\bar{\theta}$ is a steady state if $\dot{\theta}\Big|_{\theta=\bar{\theta}} = 0$.

(ii) A vector $\bar{\theta}$ is a stable steady state if for any $\epsilon > 0$ there exists a $\delta > 0$ such that if $\|\theta_0 - \bar{\theta}\| \leq \delta$ then $\|\theta_t - \bar{\theta}\| \leq \epsilon$ for all $t \geq 0$.

(iii) A vector $\bar{\theta}$ is an asymptotically stable steady state if it is stable, and if for some $\eta > 0$, if $\|\theta_0 - \bar{\theta}\| \leq \eta$ then $\lim_{t\to\infty} \theta_t = \bar{\theta}$.

Example 2.7 *Consider the replicator dynamics for the Doves-Hawks game with the fitness matrix shown. Think of the doves and hawks as two types of birds that need to share resources, such as food. A dove has*

		Player 2: Dove	Player 2: Hawk
Player 1	Dove	3,3	1,5
	Hawk	5,1	0,0

higher fitness, 3, against another dove than against a hawk, A hawk has a high fitness against a dove (5) but zero fitness against another hawk; perhaps the hawks fight over their resources. The two-dimensional state vector $(\theta_t(D), \theta_t(H))$ has only one degree of freedom because it is a probability vector. For brevity, let $x_t = \theta_t(D)$, so $\theta_t = (x_t, 1 - x_t)$. Observe that

$$\begin{aligned}u_t(D, \theta_t) &= 3x_t + (1 - x_t) = 2x_t + 1 \\ u_t(H, \theta_t) &= 5x_t \\ u_t(\theta_t, \theta_t) &= x_t(2x_t + 1) + (1 - x_t)(5x_t) = 6x_t - 3x_t^2\end{aligned}$$

So (2.4) *gives*

$$\begin{aligned}\dot{x}_t &= x_t(u_t(D,\theta_t) - u_t(\theta_t,\theta_t)) \\ &= x_t(3x_t^2 - 4x_t + 1) \\ &= x_t(1-x_t)(1-3x_t). \end{aligned} \tag{2.5}$$

Sketching the right-hand side of (2.5) *vs.* x_t *and indicating the direction of flow of* x_t *shows that* 0 *and* 1 *are steady states for* x_t *that are not stable, and* $\frac{1}{3}$ *is an asymptotically stable point for* x_t. *See Figure 2.1 Consequently,* $(1,0)$ *and* $(0,1)$ *are steady states for* θ_t *that are not stable, and* $(\frac{1}{3},\frac{2}{3})$ *is an asymptotically*

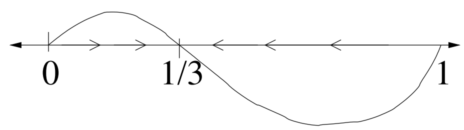


Figure 2.1: Sketch of $h(x) = x(1-x)(1-3x)$ as both a function and one-dimensional vector field.

stable point for θ_t. *In fact, if* $0 < \theta_0(D) < 1$ *then* $\theta_t \to (\frac{1}{3},\frac{2}{3})$ *as* $t \to \infty$.

Definition 2.8 *(Trembling hand perfect equilibrium) A strategy vector* $(p_1,\ldots,p_n)$ *of mixed strategies for a normal form game is a trembling hand perfect equilibrium if there exists a sequence of fully mixed vectors of strategies* $p^{(n)}$ *such that* $p^{(n)} \to (p_1,\ldots,p_n)$ *and*

$$p_i \in B_i(p_{-i}^{(n)}) \text{ for all } i,n \tag{2.6}$$

Remark 2.9 *(i) The terminology "trembling hand" comes from the image of any other player* j *intending to never use an action* a *such that* $p_j(a) = 0$, *but due to some imprecision, the player uses the action with a vanishingly (as* $n \to \infty$*) small positive probability.*

(i) The definition requires 2.6 *to hold for some sequence* $p^{(n)} \to (p_1,\ldots,p_n)$, *not for every such sequence.*

(i) Trembling hand perfect equilibrium is a stronger condition than Nash equilibrium; Nash equilibrium only requires $p_i \in B_i(p_{-i})$ *for all* i.

Figure 2.2 gives a classification of stability properties of states for replicator dynamics based on a symmetric two-player matrix game. Perhaps the most interesting implication shown in Figure 2.2 is the topic of the following proposition.

Proposition 2.10 *If* $\bar{\theta}$ *is an evolutionarily stable strategy (ESS) then it is an asymptotically stable state for the replicator dynamics.*

Proof. Fix an ESS probability vector $\bar{\theta}$. We use Kulback-Leibler divergence as a Lyapunov function. In other words, define $V(\theta)$ by

$$V(\theta) = D(\bar{\theta}\|\theta) \triangleq \sum_{a:\bar{\theta}(a)>0} -\bar{\theta}(a)\ln\frac{\theta(a)}{\bar{\theta}(a)}.$$

It is well known that $D(\bar{\theta}\|\theta)$ is nonnegative, jointly convex in $(\bar{\theta},\theta)$, and $D(\bar{\theta}\|\theta) = 0$ if and only if $\theta = \bar{\theta}$. Therefore, $V(\theta) \geq 0$ with equality if and only if $\theta = \bar{\theta}$. Also, $V(\theta)$ is finite and continuously differentiable for θ sufficiently close to $\bar{\theta}$, because such condition ensures that $\theta(a) > 0$ for all a such that $\bar{\theta}(a) > 0$.

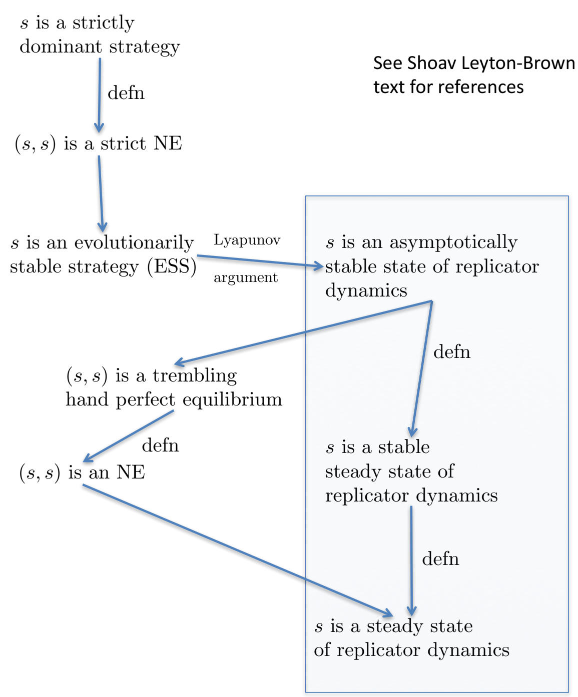


Figure 2.2: Classification of states, for evolution based on a symmetric two-person normal form game. The three definitions in the shaded box involve the replicator dynamics; the other definitions only implicitly involve dynamics.

For brevity, write $\widetilde{V}_t = V(\theta_t)$. Then by the chain rule of differentiation, the replicator fitness equation, and Proposition 2.3,

$$\begin{aligned}
\dot{\widetilde{V}} &= \nabla V(\theta_t) \cdot \dot{\theta}_t \\
&= - \sum_{a:\bar{\theta}(a)>0} \frac{\bar{\theta}(a)}{\theta_t(a)} \theta_t(a)(u(a,\theta_t) - u(\theta_t,\theta_t)) \\
&= -(u(\bar{\theta},\theta_t) - u(\theta_t,\theta_t)) \\
&< 0 \text{ for sufficiently small } \|\theta_t - \bar{\theta}\|
\end{aligned}$$

Therefore $\bar{\theta}$ is an asymptotically stable state. ■

Proposition 2.11 *[4] If p is an asymptotically stable state of the replicator dynamics, then (p, p) is an isolated trembling hand perfect equilibrium of the underlying two-player game.*

Example 2.12 *Consider the replicator dynamics for the Dove-Robin-Owl game with the fitness matrix* *shown. The owls are strictly less fit than any other species, so the only equilibrium distribution including the*

		Player 2		
		Dove	Robin	Owl
Player 1	Dove	3,3	3,3	2,2
	Robin	3,3	3,3	1,1
	Owl	2,2	1,1	0,0

owls is a pure owl population, that is distribution $(0, 0, 1)$*, is a steady state of the replicator dynamics.*

So consider the population distributions that do not include a positive fraction of owls. They are distributions of the form $(x, 1-x, 0)$*. With no owls, the doves and robin populations grow in proportion, with the fractions constant. Thus* $(x, 1-x, 0)$ *is a steady state for any* x *with* $0 \leq x \leq 1$*. In fact, it is easy to check the stronger statement:* $(x, 1-x, 0)$ *is a Nash equilibrium for any* x *with* $0 \leq x \leq 1$*.*

If there is a small positive fraction of owl's around then the dove's have a slight advantage over the robins. Thus, only $(1, 0, 0)$ *is a trembling hand perfect equilibrium distribution. The fact* $(1, 0, 0)$ *is a trembling hand perfect solution is easy to verify directly from the definition. The state* $(1, 0, 0)$ *is not asymptotically stable because it is not an isolated stable state. It follows that* $(1, 0, 0)$ *is not an ESS either, which can also be checked using Proposition 2.4.*

The only remaining question is which of the strategies of the form $(x, 1-x, 0)$*, if any, are stable steady states of the replicator dynamic. A bit of analysis shows that they are all stable steady states. Indeed, if* θ_0 *is close to* $(x, 1-x, 0)$ *and if there are initially no owls, then the initial state is a steady state so that* θ_t *is as close to* $(x, 1-x, 0)$ *as the initial state is. If there is initally a small positive fraction of owls, since the fitness of the owls is dominated by the fitnesses of the other two types, the fraction of owls converges to zero exponentially quickly. Thus, while the dove's get a small boost over the robins due to the owls, the effect converges to zero as the initial fraction of owls converges to zero.*

Additional reading

See [22, pp. 225-230] and [8, Chapter 7] for additional reading. The concept of trembling hand equilibrium plays a larger role in Chapter 4.

Chapter 3

Dynamics for Repeated Games

3.1 Iterated best response

A Nash equilibrium is a fixed point of the best response function: $s \in B(s)$. One attempt to find a Nash equilibrium is iterated best response: $s^{t+1} = B(s^t)$.

Example 3.1 *Consider iterated best response for the game shown. Beginning with $s^0 = (T, L)$ we find*

		Player 2	
		L	R
Player 1	T	1,2	3,1
	B	2,1	4,3

$s^1 = (B, L)$ $s^t = (B, R)$ for $t \geq 3$. A Nash equilibrium is reached in two iterations. Note that action B is strictly dominant for player 1, so that player 1 uses B beginning with the first iteration and does not switch again. Once player 2 plays the best response to B, no more changes occur.

Example 3.2 *Consider iterated best response for the coordination game shown. Each player has strategy*

		Player 2	
		A	B
Player 1	A	1,1	0,0
	B	0,0	1,1

set $\{A, B\}$ and they both get unit payoff if they select the same strategy, and payoff zero otherwise. Consider best response dynamics with the unfortunate initial strategy profile $s^0 = (A, B)$. Then $s^t = (B, A)$ for t odd and $s^t = (A, B)$ for t even, so the payoff to both players is zero at each step. The players are chasing after each other too quickly and are always out of sync with each other.

The next section discusses a reasonably broad class of games for which iterated best response converges if players update their responses one at a time, namely, potential games.

3.2 Potential games

(Monderer and Shapley, [15])

Let $G = (I, (S_i), (u_i))$ be an n-player normal form game and let $S = \times_{i \in I} S_i$ denote the product space of action vectors.

Definition 3.3 *A* potential function *for G is a function $\Phi : S \to \mathbb{R}$ such that*

$$u_i(x, s_{-i}) - u_i(y, s_{-i}) = \Phi(x, s_{-i}) - \Phi(y, s_{-i})$$

for all $i \in I$ and all x, y in S_i and $s_{-i} \in S_{-i}$. (Here, $S_{-i} = \times_{i' \in I \setminus \{i\}} S_{i'}$.)

If G has a potential function Φ and if the action of a single player i changes, there is no restriction on the change of payoff functions of the other players. If a game has a potential function it is easy to find one, because for any two $s, s' \in S$, the difference $\Phi(s) - \Phi(s')$ can be found by changing the coordinates of s into coordinates of s' one at a time. Any order the coordinates are changed in must give the same difference, which is why the property of having a potential function is a rather special property of a game. If Φ is a potential function, then so is $\Phi + c$ for a constant c, so in searching for a potential function we can assign an arbitrary real value to $\Phi(s)$ for one strategy profile vector $s \in S$.

Example 3.4 *Recall the prisoner's dilema. Let us seek a potential function Φ. We can set $\Phi(C, C) = 0$*

		Player 2	
		C (cooperate)	D
Player 1	C (cooperate)	1,1	-1,2
	D	2,-1	0,0

without loss of generality. If the action of either player is changed from C to D then the payoff of that player increases by one, so $\Phi(C, D) = \Phi(D, C) = 1$. Starting with (D, C), if the action of player 2 changes to D, then the payoff of player 2 again increases by one, so it must be that $\Phi(D, D) = 2$. Thus, if there is a potential function, the following table must give the function. Checking all cases we can verify that indeed, the game is a potential game and the potential function is given by

		Player 2	
		C (cooperate)	*D*
Player 1	*C (cooperate)*	$\Phi = 0$	$\Phi = 1$
	D	$\Phi = 1$	$\Phi = 2$

In a homework problem we address the following question: Does every symmetric two-player game have a potential function?

The proof of the following proposition is simple and is left to the reader.

Proposition 3.5 *Suppose G has a potential Φ.*

(i) For $s \in S$, s is a pure strategy Nash equilibrium (NE) if and only if s is a local maximum of Φ (where the neighbors of s are strategy profiles obtained by changing the strategy for a single player).

(ii) If G is a finite game (meaning that all action sets are finite) then $\arg\max \Phi \neq \emptyset$ and there exists at least one pure Nash equilibrium. Furthermore, if s^0 is an arbitrary strategy vector and a sequence $s^0, s^1, \ldots$ of strategy vectors is determined by using single player better response dynamics (players

consider whether to change strategies one at a time and any time a player can change strategies to strictly increase the payoff of the player, some such change is made), then the sequence terminates after a finite number of steps and ends at a NE.

Example 3.6 *(Single resource congestion game) Suppose there is some resource that can be shared, such as a swimming pool. Each player makes a binary decision, so that* $s_i \in S_i = \{0, 1\}$ *for each* i*, where* $s_i = 1$ *denotes that player* i *will participate in sharing the resource. The total number of players participating is* $|s| \triangleq s_1 + \ldots + s_{|I|}$*. Suppose the payoff of a player is zero if the player does not participate, and the payoff is* $v(k)$ *if a total of* k *players participate, where* $v : \{0, \ldots, |I|\} \to \mathbb{R}$*. Hence,* $u_i(s) = s_i v(|s|)$*. Thus,* $v(k)$ *represents the value of participating for a player given that a total of k players are participating. Some values of v could be negative. In many applications,* $v(k)$ *is decreasing in k, for example representing a fixed reward for participating minus a congestion cost that increases with the number of participants. Is this a potential game?*

Solution: *Let us seek a potential function* Φ*. By definition, a function* Φ *is a potential function if and only if it satisfies*

$$u_i(1, s_{-i}) - u_i(0, s_{-i}) = \Phi(1, s_{-i}) - \Phi(0, s_{-i})$$

for any player i *and any* $s_{-1} \in S_{-i}$*. If* s_{-i} *has* k *ones, this becomes*

$$v(k + 1) - 0 = \Phi(1, s_{-i}) - \Phi(0, s_{-i}).$$

That is because, if k *other players are participating, if another player decides to participate, the payoff of that player goes from 0 to* $v(k + 1)$*. Here,* $(0, s_{-i})$ *and* $(1, s_{-i})$ *can represent an arbitrary pair of strategy profiles such that the former has* k *ones, the latter has* $k + 1$ *ones, obtained by changing a 0 to a 1 in the former. Thus, it must be that* $\Phi(s) = \Phi(0) + v(1) + \cdots + v(|s|)$ *for any* $s \neq (0, \ldots, 0)$*. We can then check by similar considerations that* Φ *is indeed a potential for the game.*

As a slight generalization of this example, a player dependent participation price p_i *can be incorporated to give* $u_i(s) = s_i(v(|s|) - p_i)$*, with the corresponding potential* $\Phi(s) = v(1) + \cdots + v(|s|) - \sum_{i \in I} s_i p_i$*.*

Example 3.7 *(Multiple resource congestion game) We can extend the previous example to a set* $\mathcal{L}$ *of* $L \geq 1$ *resources. Each resource can be shared by a number of players. Each player i can select a bundle of resources* s_i *to participate in sharing, such that the bundle for player i is required to be chosen from some set of bundles* S_i*. In other words,* S_i *is a nonempty set of subsets of* $\mathcal{L}$*, and if* $s_i \in S_i$ *then* $s_i \subset \mathcal{L}$*. For example, the resources could represent communication links or transportation links in a graph, and a bundle of resources could represent a path through the graph or a set of paths through the graph. The payoff functions are given by:*

$$u_i(s_i, s_{-i}) = R_i(s_i) + \sum_{\ell \in s_i} v_\ell(k_\ell)$$

where

- $R_i : S_i \to \mathbb{R}$ *for each player* i*,*
- k_ℓ *for each* ℓ *is the number of players using resource* ℓ *(i.e. number of* j *with* $\ell \in s_j$*.)*
- $v_\ell : \{0, \ldots, |I|\} \to \mathbb{R}$ *for each* $\ell \in \mathcal{L}$*,*

The same reasoning used in Example 3.6 shows that this is a potential game for the potential function

$$\Phi(s) = \sum_i R_i(s_i) + \sum_{\ell \in \mathcal{L}} \sum_{k:1 \leq k \leq k_\ell} v_\ell(k).$$

There are not many practical examples of potential games beyond congestion games. However, Proposition 3.5 holds for games that have an ordinal potential function, defined as follows.

Definition 3.8 *(Monderer and Shapley, [15]) A function* $\Phi : S \to \mathbb{R}$ *is an* ordinal potential function *for a game* $G = (I, (S_i), (u_i))$ *if*

$$u_i(s_i', s_{-i}) > u_i(s_i, s_{-i}) \text{ if and only if } \Phi(s_i', s_{-i}) > \Phi(s_i, s_{-i})$$

for all $i \in I$ *and all* $s = (s_i, s_{-i}) \in S$ *and* $s_i' \in S_i$.

Example 3.9 *(Cournot competition) Suppose there is a single commodity, such as sugar, and a finite set of players that are firms that can produce the commodity. The action of each player* $i \in I$ *is* $q_i \in (0, \infty)$, *where* q_i *denotes the quantity of the commodity player* i *will produce. The total amount of quantity produced is given by* $Q = q_1 + \cdots + q_I$. *Suppose that the demand for the quantity is such that all the commodity will be consumed for any* $Q > 0$ *with the market price per unit quantity equal to* $P(Q)$ *for some function* $P : (0, \infty) \to (0, \infty)$. *Suppose the production cost per unit commodity is* C *for each player, where* $C > 0$. *Then the payoff function of player* i *is defined by:*

$$u(q_i, q_{-i}) = q_i(P(Q) - C)$$

So the payoff is the quantity produced by firm i *times the price per unit production minus the cost per unit production. We could make this a finite game by restricting the quantity* q_i *to be selected from some finite nonempty subset* S_i *of* $(0, \infty)$ *for each* i. *It is easy to check that the following specifies an ordinal potential function* Φ *for this game:*

$$\Phi(q) = \left(\prod_{j \in I} q_j \right) (P(Q) - C),$$

by noticing that for any i *fixed,* $\Phi(q) = \left(\prod_{j \in I \setminus \{i\}} q_j \right) u(q_i, q_{-i})$.

3.3 Fictitious play

A similar but smoother algorithm than iterated best response is fictitious play. In iterated best response, each player gives the best response to the most recent play of the other players. In fictitious play, each player gives a best response to the empirical distributions of all past plays of the other players. To describe fictitious play we use the following notation. Fix a finite n-player normal form game $(I, (S_i)_{i \in I}, (u_i)_{i \in I})$.

- Let $s^t = (s_i^t)_{i \in I}$ denote the strategy profile used at time t for integer $t \geq 1$. Although we assume the players select pure strategies at each time, we write s_i^t as a 0-1 probability vector. For example, if S_i has four elements that are indexed by 1 through 4, then $s_i^t = (0, 1, 0, 0)$ indicates that player i selects the second action from S_i at the t^{th} play of the game.

- Let $n_i^t(a)$ denote the number of times player i played action a up to time t. Set $n_i^t = (n_i^t(a))_{a\in S_i}$.
- To get an initial state, suppose t_o is a positive integer and suppose for each player i that $(n_i^0(a) : a \in S_i)$ is a vector of nonnegative integers with sum t_o. Pretend the game was played t_0 times up to time 0 and that $n_i^0(a)$ is the number of times player i selected strategy a up to time 0.
- Let $\mu_i^t = \left(\frac{n_i^t(a)}{t_o+t}\right)_{a\in S_i}$, for $t \geq 0$, which is the empirical probability distribution of the actions taken by player i up to time t, including the t_o pretend actions.
- Let $\mu^t = (\mu_1^t, \ldots, \mu_n^t)$ be the tuple of empirical probability distributions of strategies used by all players, observed in play up to time t.

Since $n_i^{t+1} = n_i^t + s_i^{t+1}$ for $t \geq 0$, we have the following update equation for the empirical distributions:

$$\mu_i^{t+1} = \frac{n_i^{t+1}}{t+t_o+1} = \frac{n_i^t + s_i^{t+1}}{t+t_o+1}$$
$$= \mu_i^t + \alpha_t(s_i^{t+1} - \mu_i^t),$$

where $\alpha_t = \frac{1}{t+t_o+1}$. The *fictitious play* algorithm is given by

$$s_i^{t+1} \in B_i(\mu_{-i}^t) \tag{3.1}$$
$$\mu_i^{t+1} = \mu_i^t + \alpha_t(s_i^{t+1} - \mu_i^t), \tag{3.2}$$

where $\alpha_t = \frac{1}{t+t_o+1}$. Equation (3.1) indicates that player i is selecting a best response at time $t+1$ as if the other players each use a mixed strategy with probability distribution equal to the empirical distribution of their past actions, and (3.2) simply updates the empirical distribution of past plays.

Proposition 3.10 *Suppose the fictitious play algorithm* (3.1) - (3.2) *is run and suppose the vector of empirical distributions converges: $\mu^t \to \sigma$ as $t \to \infty$ for some mixed strategy profile σ. Then σ is a Nash equilibrium.*

Proof. Suppose $\mu^t \to \sigma$ as $t \to \infty$. Then σ is a mixed strategy profile for the game because pointwise limits of probability vectors over finite action sets are probability vectors. It remains to show that σ is a Nash equilibrium. Focus on any player i and let $a \in S_i$ such that $a \notin B_i(\sigma_{-i})$. It suffices to show that $\sigma_i(a) = 0$. By the assumption $\mu^t \to \sigma$ it follows that $a \notin B(\mu_{-i}^t)$ for all sufficiently large t. Therefore, $\mu_i^t(a) = 0$ for all sufficiently large t, which implies $n_i^t(a)$ stops increasing for large enough t. Therefore, $\sigma_i(a) = \lim_{t\to\infty} \mu_i^t(a) = 0$. ■

So the good news about fictitious play is that if the empirical distribution converges then the limit is a Nash equilibrium. The bad news is that the empirical distribution might not converge, and even if it does, the Nash equilibrium found might not be a very desirable one.

Example 3.11 *(Fictitious play for the coordination game) Consider the coordination game of Example 3.2. To be definite, suppose in the case of a tie for best response actions, a best response action is selected at random with each possibility having positive probability bounded away from zero for all time. Consider the reasonable initial state $t_o = 2$ and $n_i^0 = (1,1)$ for both players. In other words, pretend that up to time zero each player selected each action one time. Then for the initial selection at time 1, either response for* *either player i is a best response to μ^0_{-i}. If the players are lucky and they both select the same strategy, then they will forever continue to select the same strategy, and their payoffs will both be one for every play. Moreover, the empirical distribution will converge to one of the two Nash equilibria in pure strategies, corresponding to (A, A) or (B, B). If they are unlucky at $t = 1$ and player 1 selects A and player 2 selects B, then $\mu^1 = ((2/3, 1/3), (1/3, 2/3))$. So for $t = 2$, player 1 selects the best response for distribution (1/3,2/3) for player 2, which is action B, or as a mixed strategy, action $(0, 1)$. Similarly, player 2 selections action A, or $(1, 0)$ and the payoffs are both zero at tine $t = 2$. Then the empirical distribution after two steps is $\mu^2 = ((2/4, 2/4), (2/4, 2/4))$, which is again a tie. Whenever there is such a tie the two players could again be unlucky and suffer two more rounds of zero payoff. However, with probability one they will eventually select the same actions when in a tied state, and from then on they will receive payoff 1 every time and the empirical distribution will converge to one of the two pure strategy equilibriums.*

However, consider a variation such that in case of a tie, player 1 selects action A, and in case of a tie, player 2 selects action B. Then from the same initial state as before the sequence of strategy profiles will be $s^t = (A, B)$ for t odd and $s^t = (B, A)$ for t even, and the payoffs of both players are zero for all plays. The empirical distributions of plays converges for both players, with the limit of empirical distribution profile given by $\lim_{t\to\infty} \mu^t = \sigma = ((0.5, 0.5), (0.5, 0.5))$. Notice that σ is indeed a mixed strategy Nash equilibrium as guaranteed by Proposition 3.10, but the (expected) payoff of each player is only 0.5 for each time.

Example 3.12 *(Fictitious play for Shapley's modified rock scissors paper) Consider the variation of rock scissors paper such that the payoff of the loser is 0 instead of -1, giving the following payoff matrix: This*

		Player 2		
		R	S	P
	R	0,0	1,0	0,1
Player 1	S	0,1	0,0	1,0
	P	1,0	0,1	0,0

is no longer a zero sum game. It has some aspect of the coordination game in it because the sum of payoffs is maximized (equal to one) if the players manage to select different actions. The algorithm is arbitrarily initialized to (R, S) for the first round of play. Player 1 continues to play R at $t = 2$ and $t = 3$ because of the initial play S by player 2. Player 2 switches to P at $t = 2$ and stays there awhile, as player 1 keeps playing rock. Eventually player 1 switches to S is response to the large string of P by player 2. Player 2 later switches to R because of the long string os S by player 1. And so forth. Each player moves from a losing action to a winning action, eventually causing the other player's action to become losing, and then the other player changes actions, and the process continues. The empirical distribution does not converge for fictitious play in this example.

The payoff sequence is good – the sum of payoffs is always the maximum possible value, namely one.

Remark 3.13 *Fictitious play is pretty difficult to analyze, but it does converge for the following special cases of two-player games: (a) zero sum games, (b) common interest games with randomization at ties, (c) games such that at least one player has at most two actions, with randomization at ties. See [21] for references.*

Remark 3.14 *The examples consider only two players. For three or more players we could have a player respond to either the strategy profile composed of the empirical distributions of each of the other players, or to the empirical distribution of the joint past plays of the other players.*

t	s_1^t	s_2^t	n_1^t	n_2^t
1	R	S	(1,0,0)	(0,1,0)
2	R	P	(2,0,0)	(0,1,1)
3	R	P	(3,0,0)	(0,1,2)
4	S	P	(3,1,0)	(0,1,3)
5	S	P	(3,2,0)	(0,1,4)
6	S	P	(3,3,0)	(0,1,5)
7	S	P	(3,4,0)	(0,1,6)
8	S	R	(3,5,0)	(1,1,6)
9	S	R	(3,6,0)	(2,1,6)
10	S	R	(3,7,0)	(3,1,6)
11	S	R	(3,8,0)	(4,1,6)
12	S	R	(3,9,0)	(5,1,6)
13	S	R	(3,10,0)	(6,1,6)
14	S	R	(3,11,0)	(7,1,6)
15	P	R	(3,11,1)	(8,1,6)
⋮	⋮	⋮	⋮	⋮

3.4 Regularized fictitious play and ode analysis

The fictitious play algorithm given by (3.1) and (3.2) is difficult to analyze because the best response set $B_i(\mu_{-i}^t)$ can have more than one element, so we have a difference inclusion rather than a difference equation. Also, the set $B_i(\mu_{-i}^t)$ is not continuous in μ_{-i}^t. We use a regularization technique to perturb the dynamics to address both of those problems. This section is largely based on [21]. Probability distributions are taken to be column vectors.

3.4.1 A bit of technical background

The *entropy* of a discrete probability distribution p is $H(p) \triangleq \sum_{i=1}^n -p_i \ln p_i$, with the convention that $0 \ln 0 = 0$. The entropy of a distribution on a finite set is bounded: $0 \leq H(p) \leq \ln n$. Larger values of entropy indicate the distribution is more spread out. $H(p) = 0$ if and only if p is a degenerate probability distribution, concentrated on a single outcome. The mapping $p \mapsto H(p)$ is concave.

The *Kullback-Liebler (KL) divergence* between discrete probability distributions p and q is defined by

$$D(p\|q) = \sum_i p_i \ln \frac{p_i}{q_i},$$

with the convention that $0 \ln 0 = 0 \ln \frac{0}{0} = 0$. Values of the KL divergence range over $0 \leq D(p\|q) \leq +\infty$, with $D(p\|q) = 0$ if and only if $p = q$. The mapping $(p,q) \mapsto D(p\|q)$ is jointly convex.

Let $\sigma : \mathbb{R}^m \to \mathbb{R}$ be defined by $\sigma(r) = \sigma(r_1, \ldots, r_m) = \left(\frac{e^{r_1}}{e^{r_1}+\cdots+e^{r_m}}, \cdots, \frac{e^{r_m}}{e^{r_1}+\cdots+e^{r_m}} \right)^T$. Also, define $F : \mathbb{R}^m \to \mathbb{R}$ by $F(r) = \tau \ln \left(\sum_{a=1}^m e^{r_a/\tau} \right)$, where $\tau > 0$ is a regularization parameter. F satisfies $\lim_{\tau \to 0} F(r) = \max_a r_a$ for each r and $\nabla F(r) = \sigma(r/\tau)$. For small $\tau > 0$, $F(r)$ approximates the maximum of r, and $\sigma(r/\tau)$ approximates the probability distribution uniformly distributed over the set of indices a that maximize r_a.

3.4.2 Regularized fictitious play for one player

Consider the special case of fictitious play as described in Section 3.3 for a game with only one player, namely, player 1. Equivalently, there could be other players in the game, but they always play the same actions, so the payoff function for player 1 in each round is a time invariant function of the action chosen by player 1. Write the payoff function for player 1 as $u_1(a) = r_a$ for $a \in S_1$, so r_a is the payoff, or reward, to player 1 for selecting action a. View r as a column vector. The set of mixed strategy best responses for player 1 is $B_1 = \arg\max_{p\in\Sigma_1} p^T r$, where Σ_1 is the set of probability distributions on S_1. The set B_1 does not depend on time because the payoff function does not depend on time. It is equal to the set of probability distributions over S_1 supported on the subset $\arg\max_a r_a$. The set B_1 can have more than one element and we didn't specify a probability distribution over B_1, and the set B_1 is not a continuous function of r (which would be a problem if r is changing with time, as it will in the case of more than one player).

To obtain more tractable performance consider the following regularized payoff for player 1: $u_1(p) = p^T r + \tau H(p)$, where $H(p)$ is the entropy of p and $\tau > 0$ is a regularization parameter. The best response for player 1 in mixed strategies is now uniquely given by the following continuous function of r:

$$\beta(r) \triangleq \arg\max_p \left\{p^T r + \tau H(p)\right\} = \sigma\left(\frac{r}{\tau}\right),$$

and for that distribution, the maximum payoff is $F(r) = \tau \ln \sum_a e^{r_a/\tau}$, where σ and F are defined in Section 3.4.1.[1] Using this regularization, the best response dynamics of (3.1) and (3.2) for player 1 reduces to the following discrete time randomized algorithm to be executed at each $t \geq 0$:

$$\text{select } A^{t+1} \in S_1 \text{ with probability distribution } \beta(r) \tag{3.3}$$

$$\mu_1^{t+1} = \mu_1^t + \alpha_t(\mathbf{1}_{\{A_{t+1}\}} - \mu_1^t) \tag{3.4}$$

where $\alpha_t = \frac{1}{t+t_o+1}$, and $\mathbf{1}_{\{A_{t+1}\}}$ denotes the probability distribution on S_1 with unit mass at A_{t+1}.

In the remainder of this section we shall show in a round about way that $\mu_1^t \to \beta(r)$ a.s. as $t \to \infty$. This result follows immediately from the strong law of large numbers, but the point is that the proof we give can be extended to the case r is slowly time varying. When player 1 is the only player, the empirical distribution μ_1^t is not relevant to the actions taken by the player, because we assume the player repeatedly uses the mixed strategy $\beta(r)$. But if there were additional players involved using fictitious play, μ_1^t would influence their actions.

Equation (3.4) can be written as

$$\mu_1^{t+1} = \mu_1^t + \alpha_t(\beta(r) - \mu_1^t + D_{t+1}) \tag{3.5}$$

where $D_{t+1} = \mathbf{1}_{\{A_{t+1}\}} - \beta(r)$. By (3.3), the random vector D_{t+1} is mean zero. It is also implicit in (3.3) that the selection of A_{t+1} is made independently of all past actions, so the random variables $(D_t)_{t\geq 1}$ are mutually independent, mean zero, and they are also bounded. In order to use a tool that generalizes to the case that r is time varying as in the next section when another player controls it, we use a weaker property of $(D_t)_{t\geq 1}$. Namely, $(D_t)_{t\geq 0}$ is a bounded martingale difference sequence. By the theory of stochastic approximation, it follows from (3.5) that any limit point of (μ_1^t) is in the set of fixed points for the ordinary differential equation (ode):

$$\dot{\mu_1} = \alpha_t(\beta(r) - \mu_1). \tag{3.6}$$

[1] In the terminology of statistical mechanics, if we take $-r_a$ to be the (internal) energy of a and τ to be the temperature of the system, then $p^T r + \tau H(p)$ represents the (Hemholtz) free energy of p and the probability distribution that minimizes the free energy, $\beta(r)$. is known as the Gibbs probability distribution for energy function $(-r_a)_{a\in S_1}$ and temperature τ.

The factor α_t in the ode (3.6) can be eliminated by nonlinearly rescaling time. Specifically, let q be a solution to the following ode:

$$\dot{q}(t) = \beta(r) - q(t), \tag{3.7}$$

Then $\mu_1(t) = q\left(\int_0^t \alpha_s ds\right)$ is a solution to (3.6). In summary, (3.7) represents the ode approximation of the dynamics of the empirical distribution of player 1 under fictitious, with time rescaling to eliminate α_t. We turn next to the analysis of (3.7).

The ode (3.7) is a simple linear ode, with solution given by $q(t) = \beta(r) + (q(0) - \beta(r))e^{-t}$. This directly proves $\lim_{t\to\infty} q(t) = \beta(r)$. Let's prove the same thing without solving for q, using the Lyapunov function $V(q) = F(r) - u_1(q) = F(r) - q^T r - \tau H(q)$. This choice of V is motivated by the fact $u_1(q)$ is the payoff function player 1 seeks to maximize, and $V(q)$ is obtained by subtracting the payoff from the maximum possible payoff, $F(r)$, which is achieved uniquely by $q = \beta(r)$. Alternatively, V can be expressed as $V(q) = \tau D(q\|\beta(r))$, where $D(\cdot\|\cdot)$ is the KL divergence function. In particular, $V(q) \geq 0$ and $V(q) = 0$ if and only if $q = \beta(r)$. Check that

$$\nabla V(q) = -r + \tau \begin{pmatrix} 1 + \ln q_1 \\ \vdots \\ 1 + \ln q_n \end{pmatrix}.$$

Let $\widetilde{V}_t = V(q(t))$. Check that $\dot{\widetilde{V}}_t = -\widetilde{V}_t - \tau D(\beta(r)\|q(t))$. Therefore, $\dot{\widetilde{V}}_t \leq -\widetilde{V}_t$. Thus, $\widetilde{V}_t \leq \widetilde{V}_0 \mathrm{e}^{-t}$, implying that $\lim_{t\to\infty} q(t) = \beta(r)$.

3.4.3 Regularized fictitious play for two players

Consider the two-player game in mixed strategies with payoff functions:

$$u_i(q_i, q_{-i}) = q_i^T M_i q_{-i} + \tau H(q_i)$$

If $\tau = 0$ this game corresponds to the game in mixed strategies for the bimatrix game with payoff matrix M_i for each player i. We consider $\tau > 0$ that is small, acting as a regularizer.

The situation faced by each player i is similar to that faced by the single player in Section 3.4.2, but with r replaced by $M_i q_{-i}$, which is a vector that can be time varying due to variation of the strategy of the other player. In particular, the best response for player i for a fixed distribution q_{-i} chosen by the other player is $\beta_i(q_{-i}) = \sigma\left(\frac{M_i q_{-i}}{\tau}\right)$. This gives rise to the following ode model for continuous time fictitious play:

$$\dot{q}_1(t) = \beta_1(q_2(t)) - q_1(t) \tag{3.8}$$
$$\dot{q}_2(t) = \beta_2(q_1(t)) - q_2(t). \tag{3.9}$$

Equations (3.8) and (3.9) can be derived from discrete time models using the theory of stochastic approximation in the same way (3.7) was derived.

To study convergence we consider Lyapunov functions similar to the one in Section 3.4.2.

$$V_1(q_1, q_2) = F(M_1 q_2) - u_1(q_1, q_2)$$
$$V_2(q_1, q_2) = F(M_2 q_1) - u_2(q_1, q_2)$$
$$V_{12}(q_1, q_2) = V_1(q_1, q_2) + V_2(q_1, q_2)$$

Let $\widetilde{V}_i = V_i(q_1(t), q_2(t))$ for $i \in \{1, 2\}$ and $\widetilde{V}_{12} = V_{12}(q_1(t), q_2(t))$.

Lemma 3.15

$$\dot{\widetilde{V}}_1 \leq -\widetilde{V}_1 + \dot{q}_1^T M_1 \dot{q}_2 \tag{3.10}$$

$$\dot{\widetilde{V}}_2 \leq -\widetilde{V}_2 + \dot{q}_2^T M_2 \dot{q}_1 \tag{3.11}$$

Proof. By the calculations from Section 3.4.2,

$$\nabla_{q_i} V_i(q_i(t), q_{-i}(t))^T \dot{q}_i \leq -V_i(q_i(t), q_{-i}(t))$$

By the fact $\nabla F(q) = \sigma(q/\tau)$, we find

$$\begin{aligned}\nabla_{q_{-i}} V_i(q_i(t), q_{-i}(t)) &= \nabla F(M_i q_{-i})^T M_i - \nabla_{q_{-i}} u_i(q_1, q_2) \\ &= (\beta_i(q_{-i}) - q_i)^T M_i = \dot{q}_i{}^T M_i\end{aligned}$$

so that $\nabla_{q_{-i}} V_i(q_i(t), q_{-i}(t))\dot{q}_{-i} = \dot{q}_i M_i \dot{q}_{-i}$. Thus,

$$\dot{\widetilde{V}}_i = \nabla_{q_i} V_i(q_i(t), q_{-i}(t))^T \dot{q}_i + \nabla_{q_{-i}} V_i(q_i(t), q_{-i}(t))^T \dot{q}_{-i} \leq -\widetilde{V}_i + \dot{q}_i^T M_i \dot{q}_{-i},$$

as was to be proved. ■

Proposition 3.16 *If $M_2 = -M_1^T$ (zero sum game) then any limit point of (q_1, q_2) is a Nash equilibrium. Any isolated Nash equilibrium point is asymptotically stable.*

Proof. For zero sum games, $M_2 = -M_1^T$ so that $\dot{q}_1^T M_1 \dot{q}_2 + \dot{q}_2^T M_2 \dot{q}_1 \equiv 0$. Thus, with $\widetilde{V}_{12} = \widetilde{V}_1 + \widetilde{V}_2$, application of Lemma 3.15 yields $\dot{\widetilde{V}}_{12} \leq -\widetilde{V}_{12}$ so that $\widetilde{V}_{12}(t) \leq V_{12}(0)\mathrm{e}^{-t}$ so that $\widetilde{V}_{12}(t) \to 0$ as $t \to \infty$. As noted in the previous section, $V_{12}(q)$ can be expressed as

$$V_{12}(q_1, q_2) = \tau D(q_1 \| \beta_1(q_2)) + \tau D(q_2 \| \beta_1(q_1))$$

For M_1, M_2, and τ fixed, there exists $\epsilon > 0$ so that the ranges of β_1 and β_2 both consist of probability distributions p such that $p \geq \epsilon \mathbf{1}$, and the KL divergence function $D(p\|q)$ is continuous over (p, q) such that $q \geq \epsilon \mathbf{1}$. Thus, if (q_1^*, q_2^*) is a limit point of $(q_1^t, q_2^t)_{t \geq 0}$ then $D(q_1^* \| \beta_1(q_2^*)) = D(q_2^* \| \beta_1(q_1^*)) = 0$, so that $q_1^* = \beta_1(q_2^*)$ and $q_2^* = \beta_1(q_1^*)$. In other words, (q_1^*, q_2^*) is a Nash equilibrium.

If (q_1^*, q_2^*) is any isolated Nash equilibrium point, then the continuity of $V_{1,2}$ already discussed implies there exist $\epsilon_2 > \epsilon_1 > 0$ so that

$$\min\{V_{1,2}(q_1, q_2) : \|q_1 - q_1^*\|_2 + \|q_2 - q_2^*\|_2 = \epsilon_2\} > \max\{V_{1,2}(q_1, q_2) : \|q_1 - q_1^*\|_2 + \|q_2 - q_2^*\|_2 \leq \epsilon_1\}.$$

It follows that if the initial point is in the set $\|q_1 - q_1^*\|_2 + \|q_2 - q_2^*\|_2 \leq \epsilon_1$ then the trajectory converges to (q_1^*, q_2^*) without ever leaving the set. Therefore, (q_1^*, q_2^*) is asymptotically stable. ■

3.5 Prediction with Expert Advice

(Goes back to Hannan and Wald, among others. See [7])

3.5.1 Deterministic guarantees

Suppose a forecaster selects a prediction $\widehat{p}_t \in \mathcal{D}$ of some outcome $y_t \in \mathcal{Y}$ for each $t \geq 1$. For a loss function $\ell : \mathcal{D} \times \mathcal{Y} \to \mathbb{R}$, the *cumulative loss* of the forecaster up to time n is given by:

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(\widehat{p}_t, y_t)$$

The sequence $y_1, y_2, \ldots$ can be arbitrary. It might be that each outcome y_t is given by a function of past and present predictions and past outcomes, or it can be thought of as being selected arbitrarily by nature or by an adversary.

An interesting approach to this problem is based on the assumption that there are some experts, and the forecaster tries to do as well as any of the experts. Suppose there are N experts and expert i makes prediction $f_{i,t} \in \mathcal{D}$ at time t. The cumulative loss for expert i is given by $L_{i,n} \triangleq \sum_{t=1}^{n} \ell(f_{i,t}, y_t)$ for $i \in [N]$. The regret of the forecaster for not following expert i is defined by:

$$R_{i,n} \triangleq \widehat{L}_n - L_{i,n} = \sum_{t=1}^{n} \underbrace{\ell(\widehat{p}_t, y_t) - \ell(f_{i,t}, y_t)}_{r_{i,t}},$$

where the i^{th} term of the sum is the instantaneous regret, $r_{i,t}$. A reasonable goal for the forecaster is to perform, averaged over time, as well as any of the experts. The average regret per unit time converges to zero if

$$\max_{1 \leq i \leq N} \frac{1}{n} (\widehat{L}_n - L_{i,n}) \stackrel{n \to \infty}{\longrightarrow} 0.$$

A desirable property for a forecaster would be to satisfy a universal performance guarantee: the average regret per unit time converges to zero for an arbitrary choice of the y's and arbitrary sequences of the experts. We'll see such forecasters exist if $\ell(p, y)$ is a bounded function that is a convex function of p for fixed y. The following lemma is a key to the construction.

Lemma 3.17 *Suppose $\ell(p, y)$ is convex in p. For any nonzero weight vector $(w_{1,t-1}, \ldots, w_{n,t-1})$ with non-negative entries, if*

$$\widehat{p}_t = \frac{\sum_i w_{i,t-1} f_{i,t}}{D_{t-1}} \qquad \textit{where} \quad D_{t-1} = \sum_{i'} w_{i',t-1} \tag{3.12}$$

then

$$\sum_i w_{i,t-1} r_{i,t} \leq 0$$

for any y_t, where $r_{i,t} = \ell(\widehat{p}_t, y_t) - \ell(f_{i,t}, y_t)$.

Proof. For any y_t,

$$\ell(\widehat{p}_t, y_t) = \ell\left(\sum_i \frac{w_{i,t-1}}{D_{t-1}} f_{i,t}, y_t\right)$$
$$\overset{(a)}{\leq} \sum_i \frac{w_{i,t-1}}{D_{t-1}} \ell(f_{i,t}, y_t) \tag{3.13}$$

where (a) follows by Jensen's inequality. The average of a constant is the constant, so

$$\ell(\widehat{p}_t, y_t) = \sum_i \frac{w_{i,t-1}}{D_{t-1}} \ell(\widehat{p}_t, y_t). \tag{3.14}$$

Subtracting each side of (3.13) from the respective sides of (3.14) and cancelling the positive factor D_{t-1} yields the lemma. ■

Lemma 3.17 shows that if the forecaster's prediction is a convex combination of the predictions of the experts, then the forecaster's loss will be less than or equal to the convex combination of the losses of the experts, formed using the same weights for both convex combinations. For example, if the forecaster uses the same action as one particular expert, then the loss of the forecaster will be no larger than the loss of that expert. If the forecaster simply uses the unweighted average of the actions of two particular experts, then the loss of the forecaster will be less than or equal to the unweighted average of the losses of those two experts. In order to do well in the long run, the forecaster needs to put more weight on the more successful experts, which are the ones giving the largest regret. One strategy would be to put all the weight on a single leading expert, but due to the arbitrary nature of the outcome sequence (y_t), the forecaster is better off putting significant weight on experts doing nearly as well as the leading expert, as implied by the following example.

Example 3.18 *Suppose $\mathcal{D} = [0,1]$, $\mathcal{Y} = \{0,1\}$, and $\ell(p,y) = |p-y|$. Suppose $N = 2$ with $f_{1,t} = 0$ for all t and $f_{2,t} = 1$ for all t. If the forecaster always follows a leading expert, then for each t, $\widehat{p}_t \in \{0,1\}$ is one of the endpoints of $\mathcal{D}$ for each t. But then it is possible for the forecaster to be so unlucky as to have maximum loss every time. In other words, it could happen that $y_t = 1$ whenever $\widehat{p}_t = 0$ and $y_t = 0$ whenever $\widehat{p}_t = 1$, resulting in the largest possible loss for the forecaster at every time, yielding $\widehat{L}_n = n$ for all $n \geq 1$. A better strategy would be to select $\widehat{p}_t \equiv 1/2$. Then $\widehat{L}_n \leq \frac{n}{2}$ for all $n \geq 1$. How do the experts do for this example? At each time, one of the two experts is correct and the other is not, so $L_{1,n} + L_{2,n} = n$ for all $n \geq 1$, so* $\min\{L_{1,n}, L_{2,n}\} \leq \frac{n}{2}$.

Given $\eta > 0$, let $\Phi : \mathbb{R}^N \to \mathbb{R}^N$ be defined by[2]

$$\Phi(x_1, \ldots, x_N) = \frac{1}{\eta} \ln \sum_{i=1}^N \mathrm{e}^{\eta x_i}.$$

For large η, Φ is a version of soft max function; $\lim_{\eta\to\infty} \Phi(x_1, \ldots, x_N) = \max\{x_1, \ldots, x_N\}$. In general, $\max\{x_1, \ldots, x_N\} \leq \Phi(x_1, \ldots, x_N) \leq \max\{x_1, \ldots, x_N\} + \frac{\ln N}{\eta}$. The gradient of Φ is given by

$$\nabla\Phi(x) = \begin{pmatrix} \frac{\mathrm{e}^{\eta x_1}}{D(x)} \\ \vdots \\ \frac{\mathrm{e}^{\eta x_N}}{D(x)} \end{pmatrix} \tag{3.15}$$

[2] Φ and η here are the same as F and $1/\tau$ in Section 3.4.1.

where $D(x) = \sum_{i'} \mathrm{e}^{\eta x_{i'}}$. Thus, $\nabla\Phi(x)$ is the transpose of a probability vector with most of the weight on the indices of maximum value. In other words, $\nabla\Phi(x)$ is a soft max selector function. The function Φ is convex – in fact the sum of log convex functions is convex in general. A way to directly verify the convexity of Φ is to note that the Hessian matrix is given by

$$H(\Phi)\Big|_x = \left(\frac{\partial^2 H(\Phi)}{\partial x_i \partial x_j}\right)_{i,j} = \eta\left(\mathrm{diag}\left(\frac{\mathrm{e}^{\eta x_i}}{D}\right) - \left(\frac{\mathrm{e}^{\eta x_i}}{D}\frac{\mathrm{e}^{\eta x_j}}{D}\right)_{i,j}\right) \tag{3.16}$$

Therefore, the Hessian matrix is symmetric and diagonal dominant, so it is positive definite.

Since the forecaster seeks to minimize the maximum regret over the N experts, we focus on the soft maximum regret, given by $\Phi(R_{1,t}, \ldots R_{N,t})$.

Let $R_t = (R_{1,t}, \ldots, R_{N,t})$ and $r_t = (r_{1,t}, \ldots, r_{N,t})$. Note that $R_t = R_{t-1} + r_t$ for $t \geq 1$. By the intermediate value form of Taylor's theorem,

$$\begin{aligned}\Phi(R_t) &= \Phi(R_{t-1} + r_t) \\ &= \Phi(R_{t-1}) + \langle \nabla\Phi(R_{t-1}), r_t\rangle + \frac{1}{2} r_t^T H(\Phi)\Big|_\xi r_t^T\end{aligned} \tag{3.17}$$

for some $\xi \in \mathbb{R}^N$ on the line segment with endpoints R_{t-1} and R_t. The second term on the righthand side of (3.17) is a convex combination of the instantaneous regrets for different experts. In view of Lemma 3.17, this suggests that the forecaster use $(\nabla\Phi(R_{t-1}))^T$ as the weight vector on the experts, to ensure that $\langle \nabla\Phi(R_{t-1}), r_t\rangle \leq 0$. In other words, in view of (3.15), suppose the forecaster uses the strategy (3.12) for the weights:

$$w_{i,t-1} = \mathrm{e}^{\eta R_{i,t-1}} \tag{3.18}$$

Lemma 3.17 implies that no matter what the values of $f_{1,t}, \ldots, f_{N,t}$ and y_t are, $\langle \nabla\Phi(R_{t-1}), r_t\rangle = \langle \widehat{p}_t, r_t\rangle \leq 0$. Since multiplying all the weights in (3.12) by the same constant does not change $\widehat{p}_t$, the same sequence for the predictor is generated by

$$w_{i,t-1} = \mathrm{e}^{-\eta L_{i,t-1}} \tag{3.19}$$

With this choice of weights, as already mentioned, $\langle \nabla\Phi(R_{t-1}), r_t\rangle \leq 0$. So we next examine the last term on the righthand side of (3.17).

Suppose the loss function is bounded. For convenience, assume it takes values in $[0,1]$: $\ell(p,y) \in [0,1]$ for all p and y. Therefore, $|r_{i,t}| \leq 1$ for all $i \in [N]$ and $t \geq 1$. This assumption and the expression (3.16) show that

$$r_t^T H(\Phi)\Big|_\xi r_t^T \leq \eta \sum_{i=1}^N \frac{\mathrm{e}^{\eta\xi_i}}{D} r_{i,t}^2 \leq \eta \sum_{i=1}^N \frac{\mathrm{e}^{\eta\xi_i}}{D} = \eta.$$

In other words, the last term on the righthand side of (3.17) is less than or equal to $\frac{\eta}{2}$. Thus, (3.17) and the assumptions made imply $\Phi(R_n) \leq \Phi(R_{n-1}) + \frac{\eta}{2}$ for all $n \geq 1$. Since $R_0 \equiv 0$, $\Phi(R_0) = \frac{\ln N}{\eta}$. Therefore, $\max_i R_{i,n} \leq \Phi(R_n) \leq \frac{n\eta}{2} + \frac{\ln N}{\eta}$. We thus have the following proposition.

Proposition 3.19 *(Maximum regret bound for exponentially weighted forecaster, fixed time horizon) Let $\mathcal{D}$ and $\mathcal{Y}$ be nonempty sets such that $\mathcal{D}$ is convex, and let $\ell : \mathcal{D} \times \mathcal{Y} \to [0,1]$ be such that $p \mapsto \ell(p,y)$ is convex for any y fixed. For some $n \geq 1$ let $((f_{i,t})_{1\leq t\leq n})_{i\in[N]}$ represent arbitrary strategies of N experts* *and let* $(y_t)_{1\le t\le n}$ *be an arbitrary outcome sequence. If the forecaster uses the weighted predictor* (3.12) *with exponential weights* $w_{i,t-1} = \mathrm{e}^{-\eta L_{i,t-1}}$ *for some* $\eta > 0$, *then*

$$\max_i R_{i,n} \le \frac{\ln N}{\eta} + \frac{n\eta}{2}. \tag{3.20}$$

In particular, if $\eta = \sqrt{\frac{2\ln N}{n}}$ *(i.e. the same* η *is used in the exponential weighting for* $1 \le t \le n$*), then*

$$\max_i R_{i,n} \le \sqrt{2n\ln N}.$$

In other words, the maximum regret grows with n *at most as the square root of the number of plays times the log of the number of experts.*

Remark 3.20 *The above development can be carried out for different potential functions, such as* $\Phi(R_t) = \left(\sum_i (R_{i,t})_+^p\right)^{\frac{1}{p}}$ *for a fixed* p *with* $p \ge 2$, *which gives rise to polynomially weighted average forecasters. Desirable properties of a potential function* Φ *would be that knowing* $\Phi(R_n)$ *should give a good bound on* $\max_i R_{i,n}$, *and the Hessian of* Φ *should be bounded or not too large.*

Remark 3.21 *A nice feature of the exponentially weighted average forecaster used above is that the term* $\mathrm{e}^{\eta\widehat{L}_n}$ *can be factored out of* (3.18) *to give the equivalent weights* (3.19) *depending only on the actions of the experts but not on the predictions of the forecaster.*

Remark 3.22 *A fact related to the previous remark is that the bound in* (3.19) *can be tightened by applying the same bounding procedure to* $\Phi(-L_t)$, *where* $L_t = (L_{t,i})_{i\in[N]}$ *and using Hoeffding's Lemma (Lemma 3.37), to yield that* (3.20) *holds with 2 replaced by 8. (Roughly speaking, the fact* $\ell(f_{t,i}, y_t) \in [0,1]$ *gives a bound smaller by a factor of 4 than using the fact* $r_{i,t} = \ell(\widehat{p}_t, y_t) - \ell(f_{t,i}, y_t) \in [-1,1]$. *) Then taking* $\eta = \sqrt{\frac{8\ln N}{n}}$ *gives* $\max_i R_{i,n} \le \sqrt{\frac{n\ln N}{2}}$. *See [7] for the proof.*

Remark 3.23 *A drawback of the use of the exponential weight rule used above is that the choice of* η *depends on the time horizon* n. *A homework problem addresses the doubling trick that can be used to get an upper bound that holds for a single forecaster and all* n.

For later reference we state a version of Proposition 3.19 that incorporates both the stronger bounding method mentioned in Remark 3.22 and a tighter way to get bounds holding for all time than the doubling trick mentioned in Remark 3.23. See [7, Theorem 2.3] for a proof.

Proposition 3.24 *Suppose the loss function* ℓ *is bounded and takes values in* $[0,1]$, *and is convex in its first argument. For all output sequences* (y_t), *if the forecaster uses the weighted predictor* (3.12) *with exponential weights* $w_{i,t-1} = \mathrm{e}^{-\eta_t L_{i,t-1}}$ *with time varying parameter* $\eta_t = \sqrt{8(\ln N)/t}$, *then*

$$\max_i R_{i,n} \le \sqrt{2n\ln N} + \sqrt{\frac{\ln N}{8}}$$

for all $n \ge 1$.

3.5.2 Application to games with finite action space and mixed strategies

The previous section concerns a forecaster selecting predictions from a convex set $\mathcal{D}$ to predict outcomes from a set $\mathcal{Y}$, seeking to minimize losses $\ell(\widehat{p}_t, y_t)$. In this section we consider a player in repeated plays of a normal form game taking actions from a finite set. Drop the hat in this section; let p_t denote the mixed strategy of the player at time t. In the game formulations, a player seeks to maximize his/her payoffs, but in this section we follow the flow from the previous section and think of a player seeking to minimize a loss function (just as in the case of player 1 in two-player zero sum games).

As seen in Chapter 1, if other players can take actions that depend on the action taken by a player, the player is at a large disadvantage. For example, in the matching pennies game, Example 1.12, player 1 would always lose if player 2 knew in advance what action player 1 had decided to take. However, if the player only needed to declare a mixed strategy and if the other players could not observe which pure action is randomly generated using the mixed strategy, then the player can often do much better on the average.

In the next part of this section we suppose that the player is only concerned with the sequence of average losses. Then we turn to the case the player is concerned with the actual sequence of losses that results from the randomly selected actions of the player.

Player concerned with expected losses Consider a player of a game with a finite space of N pure strategies indexed by $i \in [N]$, and a loss function $\ell(i, y_t)$ for each time t, where y_t represents the combined actions of other players, and $\ell : [N] \times \mathcal{Y} \to \mathbb{R}$. We permit the player to use mixed strategies, with p_t being a probability vector assigning probabilities to actions in $[N]$, representing the play of the player at time t. Given p_t and the outcome y_t,, the conditional expectation of the loss to the player for time t, $\ell(p_t, y_t)$, can be written as

$$\ell(p_t, y_t) = \sum_{i=1}^{N} p_{i,t}\ell(i, y_t).$$

We suppose for now the player is concerned with minimizing the conditional expected loss $\ell(p_t, y_t)$. In the terminology of economics, the player is *risk neutral* because the player is concerned only with the conditional expected loss, rather than the entire distribution of the loss.

We apply the framework of deterministic guarantees for prediction with expert advice in Section 3.5.1 by taking each of the N pure strategies to represent experts. In other words, expert i always selects action i. Let $\mathcal{D}$ denote the set of probability vectors assigning probabilities to outcomes in $[N]$. Note that $\mathcal{D}$ is a convex set and $p \mapsto \ell(p, y_t)$ is a convex (actually linear) function. The following is a corollary of the strengthened version of Proposition 3.19 mentioned in Remark 3.22.

Corollary 3.25 *Suppose $\ell(i, y) \in [0, 1]$ for all $i \in [N]$ and $y \in \mathcal{Y}$. Suppose the player uses the exponentially weighted strategy*

$$p_t = \left(\frac{\mathrm{e}^{-\eta \sum_{s=1}^{t-1} \ell(i,y_s)}}{\sum_{i'} \mathrm{e}^{-\eta \sum_{s=1}^{t-1} \ell(i',y_s)}} \right)_{i \in [N]} \tag{3.21}$$

for some $\eta > 0$ and all $t \in [n]$. Then

$$\underbrace{\sum_{t=1}^{n} \ell(p_t, y_t)}_{\bar{L}_n} - \min_i \sum_{t=1}^{n} \ell(i, y_t) \leq \frac{\ln N}{\eta} + \frac{n\eta}{8}$$

for any $y_1, \ldots, y_n$.

Note that if there are two players and y_t represents the play of the other player, the strategy in (3.21) is a regularized form of best response dynamics, because for each i, $\frac{1}{t-1} \sum_{s=1}^{t-1} \ell(i, y_s)$ is the loss for using action i against the empirical distribution of past plays of the other player.

Player concerned with actual losses In Corollary 3.25, $\bar{L}_n$ is the cumulative conditional expected loss (CCEL) of the player, and $\min_i \sum_{t=1}^{n} \ell(i, y_t)$ is the minimum CCEL, where the minimum is over fixed strategies i. Corollary 3.25 does not involve any probabilities – it gives a bound that holds for all sequences $(y_t)_{t\in[n]}$. In contrast, focus next on the actual losses, by taking into account the sequence of pure strategies $(I_t)_{t\in[n]}$ generated with distributions p_t. As usual in the theory of normal form games, assume the other players do not see I_t before taking their actions at time t. Since the future actions of the other players can depend on the random actions taken in earlier rounds, in this section we use uppercase $(Y_t)_{t\in[n]}$ to represent the sequence of actions of the other players, and model it as a random process. Assume for each t, Y_t is a function of $I_1, \ldots, I_{t-1}, Y_1, \ldots, Y_{t-1}$, and possibly some private, internal randomness available to the other players. Assume for each t, p_t is a function of $I_1, \ldots, I_{t-1}, Y_1, \ldots, Y_{t-1}$. Given the mixed strategy p_t selected by the player, the pure action I_t is generated at random with distribution p_t. Given p_t, the selection is made independently of Y_t. The actual loss incurred by the player is thus $\ell(I_t, Y_t)$. The player could have bad luck and experience a loss much greater than $\ell(p_t, Y_t)$. However, over a large number of plays, the good luck and bad luck should nearly cancel out with high probability. This notion can be made precise by the Azuma-Hoeffding inequality (see Proposition 3.40).

Letting $D_t = \ell(I_t, Y_t) - \ell(p_t, Y_t)$, note that $E[D_t|\mathcal{F}_{t-1}] = 0$ for all $t \geq 1$ where[3] $\mathcal{F}_{t-1} = \sigma(I_1, \ldots, I_{t-1}, Y_1, \ldots, Y_t)$. In other words, D is a martingale difference sequence. Also, $(p_t)_{t\in[n]}$ is a predictable process and $D_t \in [0 - \ell(p_t, Y_t), 1 - \ell(p_t, Y_t)]$ with probability one, and $(0 - \ell(p_t, Y_t))_{t\geq 1}$ and $(1 - \ell(p_t, Y_t)_{t\geq 1}$ are both predictable random processes. Thus by the Azuma-Hoeffding inequality, Proposition 3.40, for any $\gamma > 0$,

$$\mathbb{P}\left\{\sum_{t=1}^{n} D_t \geq \gamma\right\} \leq e^{-\frac{2\gamma^2}{n}}.$$

Given δ with $0 < \delta < 1$, if we set $e^{-\frac{2\gamma^2}{n}} = \delta$ and solve for γ in terms of δ we get $\gamma = \sqrt{\frac{n}{2} \ln \frac{1}{\delta}}$. So we can conclude that with probability at least $1 - \delta$,

$$\sum_{t=1}^{n} \ell(I_t, Y_t) - \sum_{t=1}^{n} \ell(p_t, Y_t) \leq \sqrt{\frac{n}{2} \ln \frac{1}{\delta}}. \tag{3.22}$$

The strengthened version of Proposition 3.19 mentioned in Remark 3.22 holds for an arbitrary choice of $(y_t)_{t\in[n]}$, so it also holds with probability one for a random choice, so with probability one:

$$\sum_{t=1}^{n} \ell(p_t, Y_t) - \min_i \sum_{t=1}^{n} \ell(i, Y_t) \leq \frac{\ln N}{\eta} + \frac{n\eta}{8}. \tag{3.23}$$

Adding the respective sides of (3.22) and (3.23) yields the following proposition.

Proposition 3.26 *Suppose $\ell(i, y) \in [0, 1]$ for $i \in [N]$ and all possible values of y. Suppose the player uses the exponentially weighted strategies*

$$p_t = \left(\frac{e^{-\eta \sum_{s=1}^{t-1} \ell(i, Y_s)}}{\sum_{i'} e^{-\eta \sum_{s=1}^{t-1} \ell(i', Y_s)}}\right)_{1\leq i\leq N}$$

[3]For a random vector Z, $\sigma(Z)$ represents the smallest σ-algebra of subsets such that Z is $\sigma(Z)$ measurable, meaning it is a σ-algebra and contains sets of the form $\{Z \leq c\}$ for any vector c with the same dimension as Z. In this context, $\sigma(Z)$ represents information of knowing Z. If X is a random variable, conditional expectations given $\sigma(Z)$ is equivalent to conditional expectations given Z. In other words, $E[X|Z] = E[X|\sigma(Z)]$. Both have the form $g(Z)$ for a suitable function g.

for some $\eta > 0$ *and* $t \in [n]$. *Let* δ *be a constant with* $0 < \delta < 1$. *With probability at least* $1 - \delta$,

$$\sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{i \in [N]} \sum_{t=1}^{n} \ell(i, Y_t) \leq \frac{\ln N}{\eta} + \frac{n\eta}{8} + \sqrt{\frac{n}{2} \ln \frac{1}{\delta}}. \tag{3.24}$$

Hannan consistent forecasters Proposition 3.26 leads us to the following definition for a forecaster.

Definition 3.27 *A forecaster producing the sequence* $(I_1, I_2, \ldots)$ *is* Hannan consistent *if, for any possibly randomized mechanism for generating* $Y_t = g_t(I_1, \ldots, I_{t-1}, Y_1, \ldots, Y_{t-1}, \xi_t)$ *(where the purpose of* ξ_t *is to allow randomness that is independent of* $(I_1, \ldots, I_t, Y_1, \ldots, Y_{t-1})$ *)*

$$\limsup_{n \to \infty} \frac{1}{n} \left(\sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{1 \leq i \leq N} \sum_{t=1}^{n} \ell(i, Y_t) \right) \leq 0 \ \ a.s.$$

(As usual, "a.s." stands for "almost surely" which means with probability one.)

Naturally a definition similar to Definition 3.27 would apply to a forecaster that is trying to maximize the terms $\ell(I_t, Y_t)$. Rather than $\limsup_{n \to \infty} \frac{1}{n}(\cdot) \leq 0$ we would require $\liminf_{n \to \infty} \frac{1}{n}(\cdot) \geq 0$ in that case.

Corollary 3.28 *Suppose the loss function* ℓ *is bounded and takes values in* $[0, 1]$, *and is convex in its first argument. Suppose the player uses the weighted predictor* (3.12) *with exponential weights* $w_{i,t-1} = e^{-\eta_t L_{i,t-1}}$ *with time varying parameter* $\eta_t = \sqrt{8(\ln N)/t}$. *Then the strategy is Hannan consistent.*

The proof of Corollary 3.28 is similar to the proof of Proposition 3.26, which bounds the maximum regret by the sum of two terms. The first term is from the deterministic bounds for regret based on expected loss, and the second is from the Azuma-Hoeffding inequality controlling the difference between the losses for the actual random actions I_t and the expected losses give the mixed strategies, $\ell(I_t, Y_t) - \ell(p_t, Y_t)$. Bounds for the expected losses that hold for all n are provided by Proposition 3.24, which uses the same weighted exponential rule as the corollary. Bounds for the random differences between actual and expected losses can be bounded with probability one as $n \to \infty$ by combining the Azuma-Hoeffding bound with the Borel-Cantelli lemma. See homework problem.

3.5.3 Hannan consistent strategies in repeated two-player, zero sum games

Two-player zero-sum games are described in Section 1.6. Let us consider the implication of Hannan consistent forecasters (which we call Hannan consistent strategies in this context) for repeated play of such games. Suppose $\ell = (\ell(i, j))$ is a function of actions i and j, each from some finite action space. Suppose player 1 wants to select i to minimize $\ell(i, j)$ and player 2 wants to select j to maximize $\ell(i, j)$. Let V denote the value of the game:

$$V = \min_p \max_q \ell(p, q) = \max_q \min_p \ell(p, q),$$

where p and q range over mixed strategies for the two players.

After n rounds of the game, player 1 has played $(I_1, \ldots, I_n)$ and player 2 has played $(J_1, \ldots, J_n)$. Let $\widehat{p}_n^I$ denote the empirical distribution of the plays of player 1 up to time n: $\widehat{p}_{n,i}^I = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{I_t = i\}}$. Similarly, let $\widehat{q}_n^J$ denote the empirical distribution of the plays of player 2 up to time n.

Proposition 3.29 *Consider a two-player repeated zero-sum game, where* $\ell : S_1 \times S_2 \to \mathbb{R}$ *serves as the loss function for player 1 and the payoff function for player 2, and the action spaces* S_1 *and* S_2 *are finite.*

(i) If player 1 uses a Hannan consistent strategy,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) \le V \;\; a.s. \tag{3.25}$$

If player 2 uses a Hannan consistent strategy,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) \ge V \;\; a.s. \tag{3.26}$$

If both players use Hannan consistent strategies,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) = V \;\; a.s. \tag{3.27}$$

(ii) Suppose both players use Hannan consistent strategies. If p^* is any limit point of the empirical distribution $\widehat{p}_n^I$ of player 1 then p^* is minmax optimal for player 1. Similarly, if q^* is any limit point of the empirical distribution $\widehat{q}_n^J$ of player 2 then q^* is maxmin optimal for player 2. If both p^* and q^* arise as such limit points, respectively, (p^*, q^*) is a saddle point for the game.

Proof. (i) By the definition of Hannan consistency, if player 1 uses a Hannan consistent strategy, then

$$\limsup_{n\to\infty} \left(\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) - \min_i \frac{1}{n}\sum_{t=1}^{n} \ell(i, J_t) \right) \le 0 \text{ a.s.} \tag{3.28}$$

However, for any n,

$$\min_i \frac{1}{n}\sum_{t=1}^{n} \ell(i, J_t) = \min_i \ell(i, \widehat{q}_{J,n}) \le \max_q \min_i \ell(i, q) = V \tag{3.29}$$

where $\widehat{q}_{J,n}$ is the empirical distribution of $(J_1, \ldots, J_n)$. Combining (3.28) and (3.29) implies (3.25). Similarly, (3.26) holds if player 2 uses a Hannan consistent strategy. It follows from (3.25) and (3.26) that (3.27) holds if both players use Hannan consistent strategies.

(ii) Let p^* be a limit point of $\widehat{p}_n^I$ and let $\epsilon > 0$. Then for large enough n in an appropriate subsequence:

$$\begin{aligned}
\max_j \ell(p^*, j) &\overset{(a)}{\le} \max_j \ell(\widehat{p}_n^I, j) + \epsilon \\
&= \max_j \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, j) + \epsilon \\
&\overset{(b)}{\le} \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) + 2\epsilon \\
&\overset{(c)}{\le} V + 3\epsilon
\end{aligned}$$

where (a) follows from the fact that the subsequence can be selected so that $\widehat{p}_n^I \to p^*$ as $n \to \infty$ along the subsequence, (b) follows from the fact player 2 is using a Hannan consistent strategy, and (c) follows from the fact player 1 is using a Hannan consistent strategy so that (3.25) holds. Since $\epsilon > 0$ is arbitrary, it follows that $\max_j \ell(p^*, j) \leq V$. In other words, p^* is minmax optimal for player 1. Similarly, q^* is maxmin optimal for player 2. Therefore, since two-player finite zero-sum games have no duality gap, any strategy profile (p^*, q^*) such that p^* is minmax optimal and q^* is maxmin optimal is a saddle point. ■

Remark 3.30 *1. Part (ii) is about limit points of the empirical distributions, such as $\widehat{p}_n^I$ for player 1. It does does not claim that limit points of the strategies used by player 1, $(p_t)_{t\geq 1}$, are minmax optimal.*

2. Limit points p of the joint empirical distribution are not necessarily product form. They satisfy a weaker, averaged form of correlated equilibrium:

$$\sum_i \sum_j p(i,j)\ell(i,j) \geq \sum_i \sum_j p(i,j)\ell(i',j) \text{ for all } i'$$

In contrast, the definition of correlated equilibrium would require

$$\sum_j p(i,j)\ell(i,j) \geq \sum_j p(i,j)\ell(i',j) \text{ for all } i, i'$$

However, by using enhanced versions of Hannan consistent strategies, it can be guaranteed that limit points of the joint empirical distribution are correlated equilibria.

3.6 Blackwell's approachability theorem

An elegant theorem of Blackwell includes the construction of Hannan consistent forecasters as a special case. To begin, we consider both the minmax and maxmin performance of a player participating in a repeated game. Consider, for example, the following game: We focus on player 1 and assume player 2 can select an

		Player 2: 1	2	3
Player 1	1	0,0	1,2	1,3
	2	2,1	0,0	2,3
	3	3,1	3,2	0,0

arbitrary sequence of actions, so the payoffs of player 2 are not relevant for the remainder of our discussion. We thus drop them from the matrix to get the payoffs of player 1. We shall find how well player 1 can

		Player 2: 1	2	3
Player 1	1	0	1	1
	2	2	0	2
	3	3	3	0

control the limiting average reward per play, regardless of the actions of player 2. The maxmin (mixed) strategy for player 1 is (0.0, 0.6, 0.4). Indeed, the payoff of player 1 is at least 1.2, no matter what action player 2 takes, and if player two uses the strategy (0.0, 0.4, 0.6) then the payoff of player 1 is less than or equal to 1.2 for any strategy. In other words, $((0.0, 0.6, 0.4), (0.0, 0.4, 0.6))$ is a Nash equilibrium, or saddle point pair, for player 1 seeking to maximize the payoff. So player 1 can ensure that the expected reward for any single play of the game is greater than or equal to 1.2. Thus, in repeated play, starting at any time,

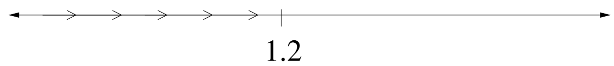


Figure 3.1: Player 1 can drive average payoff of player 1 to $[1.2, \infty)$

player 1 can ensure that the average reward per play can be pushed above any number smaller than 1.2, as illustrated in Figure 3.1. Similarly, the minmax strategy for player 1 is $(1, 0, 0)$–in other words, the pure action 1, because a Nash equilibrium, or saddle point, for the case player 1 seeks to minimize his/her return, is $((1, 0, 0), (0, b, 1 - b))$, for any b with $0 \leq b \leq 1$. The value is one. So player 1 can ensure that the reward for any single play of the game is less than or equal to 1.0. Thus, in repeated play, starting at any time, player 1 can ensure that the average reward per play can be pushed below any number greater than 1, as illustrated in Figure 3.2. Combining the above two observations, we see that for any $x_o \in [1.0, 1.2]$, player 1

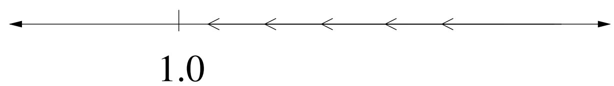


Figure 3.2: Player 1 can drive average payoff of player 1 to $(-\infty, 1]$.

can ensure that the average reward per play converges to x_o, regardless of the actions of the other player.

Similarly, let us see how well player 2 can control the limiting average reward of player 1. In view of the above two identified saddle points, we see that player 2 can ensure that the average reward per play can be pushed by player 2 to above any point less than 1.0, and can be pushed by player 2 to below any point greater than 1.2. In other words, player 2 can ensure the drift of the limiting average reward illustrated as in Figure 3.3. Define the distance between a point x and a set $S \subset \mathbb{R}$ by $d(x, S) = \inf_{s \in S} |x - s|$.

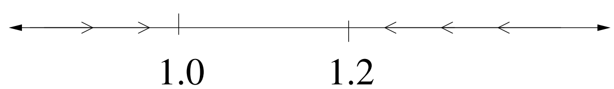


Figure 3.3: Player 2 can drive average payoff of player 1 to $[1, 2]$, but not to any specific point within $[1, 2]$

Definition 3.31 *A set $S \subset \mathbb{R}$ is* approachable *for a player if the player can play so that $d\left(\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t), S\right) \to 0$ a.s, regardless of the actions of the other player. ("a.s." means "almost surely" or "with probability one.")*

For the example above, if $A \subset \mathbb{R}$ and A is a nonempty closed subset, then

- A is a approachable for player 1 if and only if $A \cap [1, 1.2] \neq \emptyset$.
- A is a approachable for player 2 if and only if $[1, 1.2] \subset A$.

In particular,

- $[v, +\infty)$ is a approachable for player 1 if and only if $v \leq 1.2$.
- $(-\infty, u]$ is a approachable for player 1 if and only if $u \geq 1.0$.
- $[v, +\infty)$ is a approachable for player 2 if and only if $v \leq 1.0$.
- $(-\infty, u]$ is a approachable for player 2 if and only if $u \geq 1.2$.

For this example, player 1 has more control than player 2, over the limiting average of the rewards to player 1 per unit time.

The analysis of the above example can be generalized for any payoff matrix for player 1. Blackwell's approachability theorem generalizes this analysis to the case the payoff function to player 1 is vector valued. The coordinates of ℓ could represented a variety of quantities of interest to player 1. Hence, we now assume $\ell(i,j)$ takes values in $\mathbb{R}^m$ for some $m \geq 1$. Furthermore, it is assumed that ℓ is scaled, if necessary, so that $\|\ell\| \leq 1$, where "$\| \ \|$" denotes the Euclidean, or L^2, norm. Using Euclidean distance, the above definition of approachability generalizes. For $x \in \mathbb{R}^m$ and $S \subset \mathbb{R}^m$, $d(x,S) = \inf_{s\in S} \|x - s\|$. We focus on approachability by player 1.

Definition 3.32 *A set $S \subset \mathbb{R}^m$ is* approachable *for player 1 if the player can play to ensure $d\left(\frac{1}{n}\sum_{t=1}^n \ell(I_t,J_t), S\right) \to 0$ a.s., regardless of the actions of the other player.*

Lemma 3.33 *A halfspace $\mathcal{H} = \{v : u \cdot v \leq c\}$ (for $u \in \mathbb{R}^m$ and $c \in \mathbb{R}$ fixed) is approachable for player 1 if and only if for some mixed strategy p for player 1,*

$$\max_j u \cdot \ell(p,j) \leq c. \tag{3.30}$$

Proof. For u fixed, the function $(i,j) \mapsto u \cdot \ell(i,j)$ can be viewed as a scaler-valued loss function for player 1. For the two-person zero-sum game such that player 1 seeks to minimize this loss and player 2 seeks to maximize it, the value of the game is $V = \min_p \max_j u \cdot \ell(p,j)$. Player 1 can ensure that the average loss per play is eventually smaller than $c + \epsilon$ for any $\epsilon > 0$ if and only if $V \leq c$. The condition $V \leq c$ is equivalent to the existence of a mixed strategy p for player 1 such that $\max_j u \cdot \ell(p,j) \leq c$, which completes the proof. ■

Theorem 3.34 *(Blackwell approachability theorem) A closed, convex set $S \subset \mathbb{R}^m$ is approachable if and only if every halfspace containing S is approachable.*

Proof. (only if) If S is approachable then by definition, any set containing S is a also approachable, including any halfspace that contains S.

(if) The proof is illustrated in Figure 3.4. Suppose every halfspace containing S is approachable. Let $\pi_S : \mathbb{R}^m \to S$ denote the projection mapping. Thus $d(x,S) = \|x - \pi_S(x)\|$ for all $x \in \mathbb{R}^m$. We don't assume S is a subset of the unit ball, so π_S can map points in the unit ball out of the ball. However, if $\|x\| \leq 1$, then $\|\pi_S(x)\| \leq \|x\| + \|\pi_S(x) - x\| \leq 1 + d(x,S) \leq 2 + d(0,S) \triangleq M$. Also, $\|\pi_S(x) - x\| \leq M$.

Let A_0 be the zero vector in $\mathbb{R}^m$ and for $t \geq 1$ let $A_t = \frac{1}{t}\sum_{s=1}^t \ell(I_s,J_s)$. Since A_t is the average of vectors in the unit ball of $\mathbb{R}^m$, it follows that $\|A_t\| \leq 1$ for all t.

Let $t \geq 1$. By assumption, every halfspace containing S is approachable. So the smallest halfspace containing S with outgoing normal $A_{t-1} - \pi_S(A_{t-1})$ is approachable. Such halfspace can be written as $\mathcal{H} = \{x : (x - \pi_S(A_{t-1})) \cdot (A_{t-1} - \pi_S(A_{t-1})) \leq 0\}$. By Lemma 3.33 there exits a mixed strategy p such that $(\ell(p,j) - \pi_S(A_{t-1})) \cdot (A_{t-1} - \pi_S(A_{t-1})) \leq 0$ for all j. Let

$$\widehat{p}_t \in \arg\min_p \max_j \ell(p,j) \cdot (A_{t-1} - \pi_S(A_{t-1})).$$

It follows that $(\ell(\widehat{p}_t,j) - \pi_S(A_{t-1})) \cdot (A_{t-1} - \pi_S(A_{t-1})) \leq 0$ for all j.

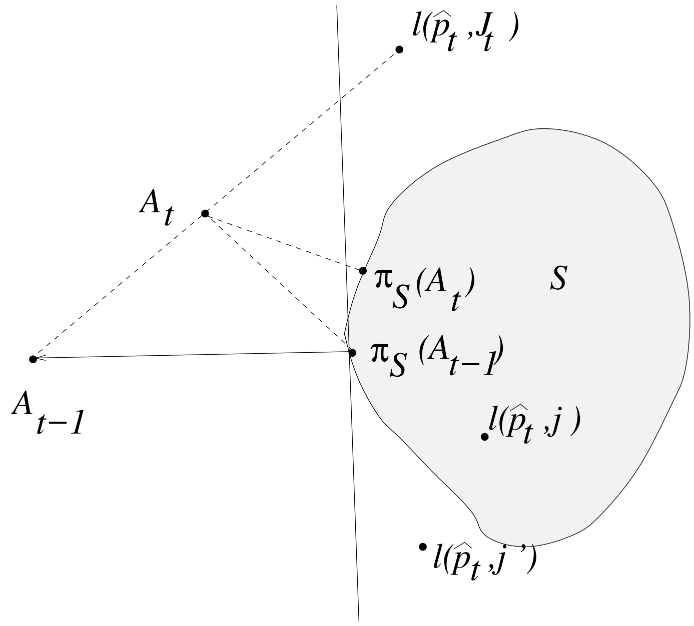


Figure 3.4: Illustration of Blackwell's approachability algorithm.

Observe that

$$\begin{aligned}
\|A_t - \pi_S(A_t)\|^2 &\le \|A_t - \pi_S(A_{t-1})\|^2 \\
&= \left\| \frac{t-1}{t}(A_{t-1} - \pi_S(A_{t-1})) + \frac{\ell(I_t, J_t) - \pi_S(A_{t-1})}{t} \right\|^2 \\
&= \left(\frac{t-1}{t}\right)^2 \|A_{t-1} - \pi_S(A_{t-1})\|^2 + \frac{1}{t^2}\|\ell(I_t, J_t) - \pi_S(A_{t-1})\|^2 \\
&\quad + \frac{2(t-1)}{t^2}\,(A_{t-1} - \pi_S(A_{t-1})) \cdot (\ell(I_t, J_t) - \pi_S(A_{t-1}))
\end{aligned} \tag{3.31}$$

The second term in (3.31) is bounded by $\frac{(M+1)^2}{t^2}$ because $\|\ell(I_t, J_t)\| \le 1$ and $\|\pi_S(A_{t-1})\| \le M$. As noted above, the last term of (3.31) would be less than or equal to zero for any value of J_t if I_t were replaced by its conditional distribution, $\widehat{p}_t$. Therefore, letting

$$D_t = \ell(I_t, J_t) - \ell(\widehat{p}_t, J_t)$$

and multiplying through by t^2 we find

$$t^2\|A_t - \pi_S(A_t)\|^2 \le (t-1)^2\|A_{t-1} - \pi_S(A_{t-1})\|^2 + (M+1)^2 + X_t$$

where

$$X_t = 2(t-1)\,(A_{t-1} - \pi_S(A_{t-1})) \cdot D_t$$

Summing over t and cancelling terms yields

$$\begin{aligned} n^2\|A_n - \pi_S(A_n)\|^2 &\le n(M+1)^2 + \sum_{t=1}^n X_t \\ \|A_n - \pi_S(A_n)\|^2 &\le \frac{(M+1)^2}{n} + \frac{1}{n^2}\sum_{t=1}^n X_t \\ &\to 0 \ \text{ a.s.} \end{aligned}$$

where the last step follows from an application of the Azuma-Hoeffding inequality and Borel-Cantelli lemma. Indeed, (X_t) is a martingale difference sequence. Also, $\|D_t\| \le 2$ and $\|A_{t-1} - \pi_S(A_{t-1})\| \le M$, so by the Cauchy-Schwarz inequality, $|X_t| \le 4(t-1)M$. Therefore, for any $c > 0$, by the Azuma-Hoeffding inequality (Proposition 3.40),

$$\mathbb{P}\left\{\sum_{t=1}^n X_t \ge cn^{3/2}\right\} \le \exp\left(-\frac{c^2 n^3}{2\sum_{t=1}^n (4(t-1)M)^2}\right) \le \exp\left(-\frac{3c^2}{16M^2}\right),$$

where we used the fact $\sum_{t=1}^n (t-1) = \frac{(n-1)n(2n-1)}{6} \le \frac{n^3}{6}$. Hence, taking $c = n^{1/4}$ yields

$$\sum_{n=1}^\infty \mathbb{P}\left\{\frac{1}{n^2}\sum_{t=1}^n X_t \ge \frac{1}{n^{1/4}}\right\} < +\infty,$$

which by the Borel-Cantelli lemma (Lemma 3.42) implies $\mathbb{P}\left\{\frac{1}{n^2}\sum_{t=1}^n X_t \ge \frac{1}{n^{1/4}} \ \text{ infinitely often}\right\} = 0$.

■

From Blackwell's theorem to Hannan consistent forecasters Suppose a forecaster selects a mixed strategy p_t which, in turn, is used to determine a pure strategy $I_t \in [N]$ with distribution p_t, at each time $t \ge 1$. Or I_t could be the index of an expert selected by the forecaster. Let $\ell : [N] \times \mathcal{Y} \to [0,1]$ such that $\ell(I_t, Y_t)$ represents the loss for using I_t at time t, assuming Y_t represents the actions of other players or choices of nature or an adversary. Define a vector valued regret function $r : [N] \times \mathcal{Y} \to [-1,1]^N$ by $r(i,y) \triangleq (\ell(i,y) - \ell(1,y), \cdots, \ell(i,y) - \ell(N,y))$. Apply Blackwell's approachability theorem to the behavior of the *time averaged regret vector,* $\frac{1}{n}\sum_{t=1}^n r(I_t, Y_t)$, as $n \to \infty$. In other words, let r play the role of the vector valued objective function in Blackwell's approachability theorem.

By definition, the forecaster is Hannan consistent if the time averaged regret vector converges to the set $S = (-\infty, 0]^N$ a.s. By Theorem 3.34, there is such a forecaster if and only if, for any nonzero vector u with nonnegative coordinates, there is a probability distribution p such that $u \cdot r(p,y) \le 0$ for all y, or in expanded form, $(u_1, \ldots, u_N) \cdot (\ell(p,y) - \ell(1,y), \cdots, \ell(p,y) - \ell(N,y)) \le 0$ for all y. It is easily checked that one such distribution is given by taking $p = u/\|u\|_1$. (This is an instance of Lemma 3.17 in case $\ell(p,y)$ is linear in p.) Thus, the existence of Hannan consistent forecasters is a corollary of Proposition 3.34.

Examining the proof of Proposition 3.34 yields the following Hannan forecaster. The strategy p_t used by the player at time t is arbitrary if the cumulative regrets satisfy $R_{i,t-1} \le 0$ for all $i \in [N]$. Otherwise, p_t is proportional to the vector $((R_{1,t-1})_+, \ldots, R_{N,t-1})_+)$. This same forecaster can be derived using the potential function of Remark 3.20 with $p = 2$.

3.7 Online convex programing and a regret bound (skip in fall 2018)

The paper of Zinkevich [24] sparked much interest in the adversarial framework for modeling online function minimization. The paper shows that a projected gradient descent algorithm achieves zero asymptotic average regret rate for minimizing an arbitrary sequence of uniformly Lipschitz convex functions over a closed bounded convex set in $\mathbb{R}^d$. The framework involves objects familiar to us, although the terminology is a bit closer to game theory.

- Let $\mathcal{F}$ be a nonempty, closed, convex subset of a Hilbert space $\mathcal{H}$. It is assumed $\mathcal{F}$ is bounded, so $D \triangleq \max\{\|f - f'\| : f, f' \in \mathcal{F}\|\} < \infty$. The player selects actions from $\mathcal{F}$.
- Let Z be a set, denoting the possible actions of the adversary.
- Let $\ell : \mathcal{F} \times \mathsf{Z} \to \mathbb{R}_+$. The interpretation is that $\ell(f_t, z_t)$ is the loss to the player for step t. We sometimes use the notation $\ell_t : \mathsf{Z} \to \mathbb{R}_+$, defined by $\ell_t(f) = \ell(f, z_t)$.
- Suppose the player has access to an algorithm that can compute $\ell_t(f)$ and $\nabla \ell_t(f)$ for a given f.
- Suppose the player has access to an algorithm that can calculate $\Pi(f)$ for any $f \in \mathbb{R}^d$, where $\Pi : \mathcal{H} \to \mathcal{F}$ is the projection mapping: $\Pi(f) = \arg\min\{\|f - f'\|^2 : f' \in \mathcal{F}\}$, that maps any $f \in \mathcal{H}$ to a nearest point in $\mathcal{F}$.
- $T \geq 1$ represents a *time horizon* of interest

The online convex optimization game proceeds as follows.

- At each time step t from 1 to T, the player chooses $f_t \in \mathcal{F}$
- The adversary chooses $z_t \in \mathsf{Z}$
- The player observes z_t and incurs the loss $\ell(f_t, z_t)$.

Roughly speaking, the player would like to select the sequence of actions (f_t) to minimize the total loss for some time-horizon T, or equivalently, minimize the corresponding average loss per time step:

$$J_T((f_t)) \triangleq \sum_{t=1}^{T} \ell(f_t, z_t.) \qquad L_T((f_t)) \triangleq \frac{1}{T} J_T((f_t)).$$

If we wanted to emphasize the dependence on z^T we could have written $J_T((f_t), z^T)$ and $L_T((f_t), z^T)$ instead. A possible strategy of the player is to use a fixed $f^* \in \mathcal{F}$ for all time, in which case we write the total loss as $J_T(f^*) \triangleq \sum_{t=1}^{T} \ell(f^*, z_t.)$ and the loss per time step as $L_T(f^*) = \frac{1}{T} J_T(f^*)$. Note that $L_T(f^*)$ is the empirical loss for f^* for T samples. If the player is extremely lucky, or if for each t a genie knowing z_t in advance reveals an optimal choice to the player, the player could use $f_t^{\text{genie}} \triangleq \arg\min_{z \in \mathsf{Z}} \ell(f, z_t)$. Typically it is unreasonable to expect a player without knowing z_t before selecting f_t to achieve, or even nearly achieve, the genie-assisted minimum loss.

It turns out that a realistic goals is for the player to make selections that perform nearly as well as *any fixed strategy* f^* that could possibly be selected after the sequence z^T is revealed. Specifically, if the player uses (f_t) then the *regret* (for not using an optimal fixed strategy) is defined by:

$$R_T((f_t)) = \inf_{f^* \in \mathcal{F}} J_T((f_t)) - J_T(f^*),$$

where for a particular f^*, $J_T((f_t)) - J_T(f^*)$ is the regret for using (f_t) instead of f^*. We shall be interested in strategies the player can use to (approximately) minimize the regret. Even this goal seems ambitious, but one important thing the player can exploit is that the player can let f_t depend on t, whereas the performance the player aspires to match is that of the best policy that is constant over all steps t.

Zinkevich [24] showed that the projected gradienct descent algorithm, defined by

$$f_{t+1} = \Pi(f_t - \alpha_t \nabla \ell_t(f_t)), \tag{3.32}$$

meets some performance guarantees for the regret minimization problem. Specifically, under convexity and the assumption that the functions ℓ_t are all L-Lipschtiz continuous, Zinkevich showed that regret $O(LD\sqrt{T})$ is achievable by gradient descent. Under such assumptions the $\sqrt{T}$ scaling is the best possible (see problem set 6). The paper of Hazan, Agarwal, and Kale [11] shows that if, in addition, the functions ℓ_t are all σ-strongly convex for some $\sigma > 0$, then gradient descent can achieve $O\left(\frac{L^2}{\sigma}\log T\right)$ regret. The paper [11] ties together several different previous approaches including follow-the-leader, exponential weighting, Cover's algorithm, and gradient descent. The following theorem combines the analysis of [24] for the case of Lipschitz continuous objective functions and the analysis of [11] for strongly convex functions. The algorithms used for the two cases differ only in the stepsize selections. Recall that D is the diameter of $\mathcal{F}$.

Theorem 3.35 *Suppose $\ell(\cdot, z)$ is convex, L-Lipschtiz continuous for each z and suppose the gradient projection algorithm* (3.37) *is run with stepsize multipliers $(\alpha_t)_{t\geq 1}$.*
(a) If $\alpha_t = \frac{c}{\sqrt{t}}$ for $t \geq 1$, then the regret is bounded as follows:

$$R_T((f_t)) \leq \frac{D^2\sqrt{T}}{2c} + \left(\sqrt{T} - \frac{1}{2}\right) L^2 c,$$

which for $c = \frac{D}{L\sqrt{2}}$ gives:

$$R_T((f_t)) \leq DL\sqrt{2T}.$$

(b) If, in addition, $\nabla\ell(\cdot, z)$ is σ-strongly convex for some $\sigma > 0$ and $\alpha_t = \frac{1}{\sigma t}$ for $t \geq 1$, then the regret is bounded as follows:

$$R_T((f_t)) \leq \frac{L^2(1 + \log T)}{2\sigma}.$$

Proof. Most of the proof is the same for parts (a) and (b), where for part (a) we simply take $\sigma = 0$. Let $f_t^\flat = f_t - \alpha_t \nabla\ell_t(f_t)$, so that $f_{t+1} = \Pi(f_{t+1}^\flat)$. Let $f^* \in \mathcal{F}$ be any fixed policy. Note that

$$\begin{aligned} f_{t+1}^\flat - f^* &= f_t - f^* - \alpha_t \nabla\ell_t(f_t) \\ \|f_{t+1}^\flat - f^*\|^2 &= \|f_t - f^*\|^2 - 2\alpha_t\langle f^t - f^*, \nabla\ell_t(f_t)\rangle + \alpha_t^2\|\nabla\ell_t(f_t)\|^2. \end{aligned}$$

By the contraction property of Π, $\|f_{t+1} - f^*\| \leq \|f_{t+1}^\flat - f^*\|$. Also, by the Lipschitz assumption, $\|\nabla\ell_t(f_t)\| \leq L$. Therefore,

$$\|f_{t+1} - f^*\|^2 \leq \|f_t - f^*\|^2 - 2\alpha_t\langle f^t - f^*, \nabla\ell_t(f_t)\rangle + \alpha_t^2 L^2$$

or, equivalently,

$$2\big\langle f_t - f^*, \nabla\ell_t(f_t)\big\rangle \leq \frac{\|f_t - f^*\|^2 - \|f_{t+1} - f^*\|^2}{\alpha_t} + \alpha_t L^2. \tag{3.33}$$

(Equation (3.33) captures well the fact that this proof is based on the use of $\|f_t - f^*\|$ as a potential function. The only property of the gradient vectors $\nabla \ell_t(f_t)$ used so far is $\|\nabla \ell_t(f_t)\| \leq L$. The specific choice of using gradient vectors is exploited next, to bound differences in the loss function.) The strong convexity of ℓ_t implies $\ell_t(f^*) - \ell_t(f_t) \geq \langle f^* - f_t, \nabla \ell_t(f_t)\rangle + \frac{\sigma}{2}\|f^* - f_t\|^2$, or equivalently,
$2(\ell_t(f_t) - \ell_t(f^*)) \leq 2\langle f_t - f^*, \nabla \ell_t(f_t)\rangle - \sigma\|f_t - f^*\|^2$, so

$$2(\ell_t(f_t) - \ell_t(f^*)) \leq \frac{\|f_t - f^*\|^2 - \|f_{t+1} - f^*\|^2}{\alpha_t} + \alpha_t L^2 - \sigma\|f_t - f^*\|^2 \tag{3.34}$$

We shall use the following for $1 \leq t \leq T-1$:

$$\frac{\|f_t - f^*\|^2 - \|f_{t+1} - f^*\|^2}{\alpha_t} = \frac{\|f_t - f^*\|^2}{\alpha_t} - \frac{\|f_{t+1} - f^*\|^2}{\alpha_{t+1}} + \left(\frac{1}{\alpha_{t+1}} - \frac{1}{\alpha_t}\right)\|f_{t+1} - f^*\|^2.$$

Summing each side of (3.34) from $t = 1$ to T yields:

$$\begin{aligned} 2(J_T((f_t)) - J_T(f^*)) &\leq \left(\frac{1}{\alpha_1} - \sigma\right)\|f_1 - f^*\|^2 - \frac{1}{\alpha_T}\|f_{T+1} - f^*\|^2 \\ &\quad + \sum_{t=1}^{T-1}\left(\frac{1}{\alpha_{t+1}} - \frac{1}{\alpha_t} - \sigma\right)\|f_{t+1} - f^*\|^2 + L^2\sum_{t=1}^{T}\alpha_t \\ &\leq D^2\left(\frac{1}{\alpha_1} - \sigma + \sum_{t=1}^{T-1}\left(\frac{1}{\alpha_{t+1}} - \frac{1}{\alpha_t} - \sigma\right)\right) + L^2\sum_{t=1}^{T}\alpha_t \\ &\leq D^2\left(\frac{1}{\alpha_T} - \sigma T\right) + L^2\sum_{t=1}^{T}\alpha_t \end{aligned} \tag{3.35}$$

(Part (a)) If $\sigma = 0$ the bound (3.35) becomes

$$2(J_T((f_t)) - J_T(f^*)) \leq \frac{D^2}{\alpha_T} + L^2\sum_{t=1}^{T}\alpha_t \tag{3.36}$$

Now if $\alpha_1 = \frac{c}{\sqrt{t}}$, then

$$\sum_{t=1}^{T}\alpha_t = c + \sum_{t=2}^{T}\frac{c}{\sqrt{t}} \leq c + c\int_{t=1}^{T}\frac{cdt}{\sqrt{t}} = (2\sqrt{T} - 1)c$$

and we get

$$J_T((f_t)) - J_T(f^*) \leq \frac{D^2\sqrt{T}}{2c} + \left(\sqrt{T} - \frac{1}{2}\right)L^2 c$$

If $c = \frac{D}{L\sqrt{2}}$ then $J_T((f_t)) - J_T(f^*) \leq DL\sqrt{2T}$. Since $f^* \in \mathcal{F}$ is arbitrary it follows that $R_T((f_t)) \leq DL\sqrt{2T}$.

(Part (b)) For the case of $\sigma > 0$ and $\alpha_t = \frac{1}{\sigma t}$ for all $t \geq 1$, then the first term on the right-hand side of (3.35) is zero, and

$$\sum_{t=1}^{T}\alpha_t = \frac{1}{\sigma}\left(1 + \sum_{t=2}^{T-1}\frac{1}{t}\right) \leq \frac{1 + \log T}{\sigma},$$

so part (b) of the theorem follows from (3.35) and the fact $f^* \in \mathcal{F}$ is arbitrary. ■

3.7.1 Application to game theory with finite action space

Consider a player repeatedly playing a game with N actions available. Let's fit this into the framework of Zinkevich [24]. We consider the case that the player uses mixed strategies, so that $\mathcal{F}$ is the space of probability vectors in $\mathbb{R}^N$. Let z_t represent the vector of actions of other players. The loss function for our player at time t is $\ell(f_t, z_t)$ where $\ell(f,z) = \sum_{i\in[N]} f_i \ell(i,z)$. The function $f \mapsto \ell(f,z)$ is linear, and thus also convex. It's gradient with respect to f is $(\ell(1,z),\ldots,\ell(N,z))^T$. We assume that $\ell(i,z)\in[0,1]$ for all i,z. Therefore, $\|\nabla_f \ell(f,z)\| \leq \sqrt{N}$. In other words, we set $L=\sqrt{N}$ in Theorem 3.35. To avoid triviality, we assume that $N\geq 2$. The difference $\|f-f'\|$ is maximized over $f,f'\in\mathcal{F}$ when f and f' correspond to different pure strategies. To verify that fact, first note that without loss of generality it can be assumed for each $i\in[N]$ that either $f_i=0$ or $f'_i=0$. Thus, we can take $D=2$ in Theorem 3.35.

The projected gradient algorithm becomes (where we view f_t as a row vector for each t) in this case is:

$$f_{t+1} = \Pi(f_t - \alpha_t(\ell(1,z),\ldots,\ell(N,z))), \tag{3.37}$$

where Π denotes the projection operator from $\mathbb{R}^N$ (with vectors written as row vectors) to $\mathcal{D}$. Interestingly, the update at time t is determined completely by the vector of losses at time t for different strategies.

The loss for T time steps for any fixed strategy f^* is given by

$$\sum_{t=1}^{T} \ell(f^*, z_t) = \langle f^*, \sum_{t=1}^{T}(\ell(1,z_t),\ldots,\ell(N,z_t))$$

so that for any sequence $z_1,\ldots,z_T$, the minimum loss over all fixed strategies is the same as the minimum loss over all pure strategies. Therefore, the regret $R_T((f_t))$ of Zinkevich is the same as the game theoretic regret. With $D=\sqrt{2}$ and $L=\sqrt{N}$ we find that for $T\geq 1$ and stepsizes as in Theorem 3.35(a), $R_T((f_t)) \leq \sqrt{4NT}$.

3.8 Appendix: Large deviations, the Azuma-Hoeffding inequality, and stochastic approximation

Recall the most basic concentration inequalities.

- Markov inequality: If Y is a random variable with $\mathbb{P}\{Y\geq 0\}=1$, then $\mathbb{P}\{Y\geq c\}\leq \frac{\mathbb{E}[Y]}{c}$ for any $c>0$, This can be proved by taking expectations on both sides of the inequality: $c\mathbf{1}_{\{Y\geq c\}}\leq Y$.

- Chebychev inequality: If X is a random variable with finite mean: $\mathbb{P}\{|X-\mathbb{E}[X]|\geq t\}\leq \frac{\mathrm{Var}(X)}{t^2}$ for any $t\geq 0$. The Chebychev inequality follows from the Markov inequality for $Y=(X-\mathbb{E}[X])^2$.

- Chernoff inequality: If $S_n = X_1+\ldots+X_n$, where the X's in independent, identically distributed with mean μ, then for any $a\geq\mu$,

$$\mathbb{P}\left\{\frac{S_n}{n}\geq a\right\}\leq e^{-n\ell(a)} \text{ where } \ell(a)=\sup_{s\in\mathbb{R}}\{as-\psi(s)\} \text{ and } \psi(s)=\log\mathbb{E}\left[e^{sX}\right].$$

  The Chernoff inequality follows from Markov's inequality with $Y=e^{-s(na-S_n)}$ and $c=1$ for $s\geq 0$. The condition $s\geq 0$ is dropped in the definition of $\ell(a)$ by the following reasoning. Note that

$$\psi'(s)=\frac{\mathbb{E}\left[Xe^{sX}\right]}{\mathbb{E}\left[e^{sX}\right]}=\mathbb{E}_s[X],$$

where $\mathbb{E}_s$ denotes expectation with respect to the new probability distribution $\mathbb{P}_s$ for X defined by $\frac{d\mathbb{P}_s}{d\mathbb{P}}(X) = \mathrm{e}^{sX-\psi(s)}$. Similarly,

$$\psi''(s) = \frac{\mathbb{E}\left[X^2\mathrm{e}^{sX}\right]}{\mathbb{E}\left[\mathrm{e}^{sX}\right]} - \mathbb{E}_s\left[X\right]^2 = \mathbb{E}_s\left[X^2\right] - \mathbb{E}_s\left[X\right]^2 = \mathrm{Var}_s\left(X\right),$$

Thus, $\psi(0) = 0$, $\psi'(0) = \mathbb{E}[X] = \mu$, and ψ is convex, because $\psi''(s) \geq 0$ for all s. These properties of ψ explain why the supremum in the definition of $\ell(a)$ can be taken over all $s \in \mathbb{R}$.

Example 3.36 *If $0 < q < p \leq 1$ and X_i has the Bernoulli distribution with parameter q, then S_n has the binomial distribution with parameter q. Then $\ell(p) = d(p||q)$ where $d(p||q) \triangleq p\log\frac{p}{q} + (1-p)\frac{1-p}{1-q}$, denotes the Kullback-Liebler (KL) divergence between the Bernoulli(p) and Bernoulli(q) distributions. So the Chernoff inequality in this case becomes*

$$\mathbb{P}\left\{Binom(n,q) \geq np\right\} \leq \mathrm{e}^{-nd(p||q)} \quad \textit{for } p \geq q > 0$$

For another example, if X_i has the $N(\mu, \sigma^2)$ distribution then $\psi(s) = \frac{s^2\sigma^2}{2} + \mu s$ and $\ell(a) = \frac{(a-\mu)^2}{2\sigma^2}$. Taking $n = 1$ in the Chernoff inequality in this case gives

$$\mathbb{P}\left\{Normal(\mu,\sigma^2) - \mu \geq t\right\} \leq \mathrm{e}^{-\frac{t^2}{2\sigma^2}}$$

In this case, we know

$$\mathbb{P}\left\{Normal(\mu,\sigma^2) - \mu \geq t\right\} = Q\left(\frac{t}{\sigma}\right) = \int_{\frac{t}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{u^2}{2}}du$$

The methodology of the Chernoff bound is quite general and flexible. In particular, if the distribution of a random variable X is not known, but $\mathbb{E}\left[\mathrm{e}^{sX}\right]$ can be bounded above, then upper bounds on large deviations can still be obtained. That is the idea of the Hoeffding inequality, based on the following lemma.

Lemma 3.37 *(Hoeffding) Suppose X is a random variable with mean zero such that $\mathbb{P}\left\{X \in [a,b]\right\} = 1$. Then $\mathbb{E}\left[\mathrm{e}^{sX}\right] \leq \mathrm{e}^{\frac{s^2(b-a)^2}{8}}$.*

Proof. Since $\mathbb{P}\left\{X \in [a,b]\right\} = 1$, it follows that $\mathbb{P}_s\{X \in [a,b]\} = 1$ for all s, and therefore, $\psi''(s) = \mathrm{Var}_s\left[X\right] \leq \frac{(b-a)^2}{4}$ for any $s \in \mathbb{R}$. Since $\psi(0) = 0$ and $\psi'(0) = \mathbb{E}[X] = 0$,

$$\psi(s) = \int_0^s \int_0^t \psi''(t)dtds \leq \frac{s^2(b-a)^2}{8},$$

which implies the lemma. ■

Note that the maximum possible variance for X under the conditions of the lemma is $\frac{(b-a)^2}{4}$, and the upper bound is equal to $\mathbb{E}\left[\mathrm{e}^{sZ}\right]$ for a $N\left(0, \frac{(b-a)^2}{4}\right)$ random variable Z.

A random process $(Y_n : n \geq 0)$ is a martingale with respect to a filtration of σ-algebras $\boldsymbol{\mathcal{F}} = (\mathcal{F}_n : n \geq 0)$ if $\mathbb{E}[Y_0]$ is finite, Y_n is $\mathcal{F}_n$ measurable for each $n \geq 0$, and $E[Y_{n+1}|\mathcal{F}_n] = Y_n$. A random process $(B_n : n \geq 1)$, is a predictable process for the filtration $\boldsymbol{\mathcal{F}}$ if B_n is $\mathcal{F}_{n-1}$ measurable for each $n \geq 1$. In other words, if B is predictable, the value B_n is determined by information available up to time $n-1$. A simple and useful inequality for martingales is the Azuma-Hoeffding inequality.

Proposition 3.38 *(Azuma-Hoeffding inequality) Let $(Y_n : n \geq 0)$ be a martingale and $(A_n : n \geq 1)$ and $(B_n : n \geq 1)$ be predictable processes, all with respect to a filtration $\boldsymbol{\mathcal{F}} = (\mathcal{F}_n : n \geq 0)$, such that $\mathbb{P}\{Y_n - Y_{n-1} \in [A_n, B_n]\} = 1$ and $\mathbb{P}\{|B_n - A_n| \leq c_n\} = 1$ for all $n \geq 1$. Then for all $\gamma \in \mathbb{R}$,*

$$\begin{aligned} P\{Y_n - Y_0 \geq \gamma\} &\leq \exp\left(-\frac{2\gamma^2}{\sum_{t=1}^n c_t^2}\right) \\ P\{Y_n - Y_0 \leq -\gamma\} &\leq \exp\left(-\frac{2\gamma^2}{\sum_{t=1}^n c_t^2}\right). \end{aligned}$$

Proof. Let $n \geq 0$. The idea is to write $Y_n = Y_n - Y_{n-1} + Y_{n-1}$, to use the tower property of conditional expectation, and to apply Lemma 3.37 to the random variable $Y_n - Y_{n-1}$ conditioned on $\mathcal{F}_{n-1}$, for $[a, b] = [A_n, B_n]$. This yields:

$$\begin{aligned} E[e^{s(Y_n - Y_0)}] &= E[E[e^{s(Y_n - Y_{n-1} + Y_{n-1} - Y_0)}|\mathcal{F}_{n-1}]] \\ &= E[e^{s(Y_{n-1} - Y_0)}E[e^{s(Y_n - Y_{n-1})}|\mathcal{F}_{n-1}]] \\ &\leq E[e^{s(Y_{n-1} - Y_0)}]e^{(sc_n)^2/8}. \end{aligned}$$

Thus, by induction on n,

$$E[e^{s(Y_n - Y_0)}] \leq e^{(s^2/8)\sum_{t=1}^n c_t^2}.$$

The remainder of the proof is the same as the proof of Chernoff's inequality for a single Gaussian random variable. ■

Corollary 3.39 *If $X_1, \ldots, X_n$ are independent random variables such that $\mathbb{P}\{X_t \in [a_t, b_t]\} = 1$ for all t, and $S_n = X_1 + \ldots + X_n$ then*

$$\begin{aligned} P\{S_n - \mathbb{E}[S_n] \geq \gamma\} &\leq \exp\left(-\frac{2\gamma^2}{\sum_{t=1}^n (b_t - a_t)^2}\right) \\ P\{S_n - \mathbb{E}[S_n] \leq -\gamma\} &\leq \exp\left(-\frac{2\gamma^2}{\sum_{t=1}^n (b_t - a_t)^2}\right). \end{aligned}$$

Proof. Apply Proposition 3.38 with $Y_0 = 0$, $Y_n = \sum_{t=1}^n (X_t - \mathbb{E}[X_t])$, $A_t = a_t$, and $B_t = b_t$ for $t \in [n]$. ■

In some applications it is more natural to consider martingale difference sequences than martingales directly. A random process $(D_1, D_2, \ldots)$ is a martingale difference sequence with respect to a filtration of σ-algebras $\boldsymbol{\mathcal{F}} = (\mathcal{F}_n : n \geq 0)$ if D_n is $\mathcal{F}_n$ measurable for each $n \geq 0$, and $E[D_{n+1}|\mathcal{F}_n] = 0$. Equivalently, $Y_0 = 0$ and $Y_n = D_1 + \cdots + D_n$ defines a martingale with respect to $\boldsymbol{\mathcal{F}}$.

Proposition 3.40 *(Azuma-Hoeffding inequality, martingale difference form) Let $(D_n : n \geq 0)$ be a martingale difference sequence and $(A_n : n \geq 1)$ and $(B_n : n \geq 1)$ be predictable processes, all with respect to a filtration $\boldsymbol{\mathcal{F}} = (\mathcal{F}_n : n \geq 0)$, such that $\mathbb{P}\{D_n \in [A_n, B_n]\} = 1$ and $\mathbb{P}\{|B_n - A_n| \leq c_n\} = 1$ for all $n \geq 1$.*

Then for all $\gamma \in \mathbb{R}$,

$$\mathbb{P}\left\{\sum_{t=1}^{n} D_t \geq \gamma\right\} \leq \exp\left(-\frac{2\gamma^2}{\sum_{t=1}^{n} c_t^2}\right)$$
$$\mathbb{P}\left\{\sum_{t=1}^{n} D_t \leq -\gamma\right\} \leq \exp\left(-\frac{2\gamma^2}{\sum_{t=1}^{n} c_t^2}\right).$$

Borel-Cantelli lemma Convergence results that hold with probability one in the limit as $n \to \infty$, such as in the definition of Hannan consistent forecasters, can be deduced from the Azuma-Hoeffding inequality by combining it with the Borel-Cantelli lemma, stated next. Let $(A_n : n \geq 0)$ be a sequence of events for a probability space $(\Omega, \mathcal{F}, P)$.

Definition 3.41 *The event* $\{A_n$ *infinitely often*$\}$ *is the set of* $\omega \in \Omega$ *such that* $\omega \in A_n$ *for infinitely many values of* n.

Another way to describe $\{A_n$ infinitely often$\}$ is that it is the set of ω such that for any k, there is an $n \geq k$ such that $\omega \in A_n$. Therefore,

$$\{A_n \text{ infinitely often}\} = \cap_{k\geq 1} \left(\cup_{n\geq k} A_n\right).$$

For each k, the set $\cup_{n\geq k} A_n$ is a countable union of events, so it is an event, and $\{A_n$ infinitely often$\}$ is an intersection of countably many such events, so that $\{A_n$ infinitely often$\}$ is also an event.

Lemma 3.42 *(Borel-Cantelli lemma) Let* $(A_n : n \geq 1)$ *be a sequence of events and let* $p_n = P(A_n)$.

(a) If $\sum_{n=1}^{\infty} p_n < \infty$, *then* $P\{A_n$ *infinitely often*$\} = 0$.

(b) If $\sum_{n=1}^{\infty} p_n = \infty$ *and* $A_1, A_2, \cdots$ *are mutually independent, then* $P\{A_n$ *infinitely often*$\} = 1$.

Stochastic approximation We close this section by stating a convergence result from the theory of stochastic approximation. The theory of stochastic approximation model was initiated by Robbins and Monroe (1951) and Kiefer and Wolfowitz (1952), and its analysis based on ordinary differential equations was initiated by Ljung (1977) and Kushner and Clark (1978). For more on this topic see [5], [3]. Suppose $(X_n)_{n\geq 0}$ is a sequence of random variables in $\mathbb{R}^d$ satisfying

$$X_{n+1} = X_n + \alpha_n \left[h(X_n) + D_{n+1} + \epsilon_{n+1}\right]$$

where

- h is Lipschitz continuous: there exists a finite constant L so that $\|h(x) - h(y)\|_2 \leq L\|x - y\|_2$ for all $x, y \in \mathbb{R}^d$.
- $(\alpha_n : n \geq 0)$ are positive constants such that $\sum_n \alpha_n = \infty$ and $\sum_n \alpha_n^2 < \infty$.
- $(D_n)_{n\geq 1}$ is a martingale difference sequence: If $\mathcal{F}_n$ represents the information generated by $(X_0, \ldots, X_n, D_1, \ldots, D_n, \epsilon_1, \ldots, \epsilon_n)$, then $\mathbb{E}\left[D_n|\mathcal{F}_n\right] = 0$ a.s., for all $n \geq 1$. In addition, $E[\|D_{n+1}\|_2^2|\mathcal{F}_n] \leq K(1 + \|X_n\|^2)$ a.s.
- $\sup_n \|X_n\|_2 < \infty$ a.s.
- The sequence ϵ_n is a deterministic or random sequence that is uniformly bounded and has limit 0.

The assumption h is Lipschitz continuous ensures that, given an initial point x_0, the following ordinary differential equation (o.d.e.) has a unique solution:

$$\dot{x}_t = h(x_t). \tag{3.38}$$

Definition 3.43 *A set $A \subset \mathbb{R}^d$ is* invariant *for the ode if $x_0 \in A$ implies $x_t \in A$ for all $t \geq 0$. A closed invariant set A is* internally chain transitive *if for any $x, y \in A$, any $\epsilon > 0$, and any $T > 0$, there is a finite sequence of points $x_0, x_1, \ldots, x_K$ in A and $t_0, \ldots, t_{K-1} \in [T, \infty)$ such that $x = x_0$, $y = x_K$, and for $0 \leq k \leq K-1$, the trajectory for the ode (3.38) with initial condition x_k is in the radius ϵ ball $B(\epsilon, x_{k+1})$ at time t_k.*

Intuitively, an internally chain transitive invariant set is a generalized limit cycle for the ode if the positions of the points are known only to accuracy ϵ for arbitrarily small ϵ. The following is stated and proved in [5], which in turn is based on [3].

Proposition 3.44 *Under the assumptions stated above, with probability one, there exits a compact, closed, connected, internally chain transitive invariant set A for the ode (3.38) such that $(X_n)_{n\geq 0}$ converges to A. In other words, $\min\{\|X_n - a\|_2 : a \in A\} \to 0$ as $n \to \infty$. (The set A can possibly be random.)*

The fact the stepsize multipliers α_n are converging to zero makes the process $(X_n)_{n\geq 0}$ move more slowly as time increases. In some application, $h = \nabla F$ for some function F that is to be maximized, and the sequence $(X_n)_{n\geq 0}$ is a noisy version of gradient ascent. If there is a globally stable point x^* for the ode (3.38)then the proposition implies $X_n \to x^*$ a.s.

Chapter 4

Sequential (Extensive Form) Games

(See [17] and [22] for more extensive treatments of this topic.)

4.1 Perfect information extensive form games

An extensive form game is specified by a rooted tree, known to each player. We begin by considering such games with *perfect information*, meaning that one player at a time makes a decision and has exact knowledge of the state of the game.

The game of Nim fits this model. Initially there are three piles of sticks, with 5 sticks in the first pile, 4 in the second pile, and 3 in the third pile. In short, the initial state is $(5, 4, 3)$. There are two players taking turns, starting with player 1. On a turn, each player must remove a nonzero number of sticks from one of the piles. If a player has picked up the last remaining stick, then the game ends and that player loses. This is the misère version of the game. An alternative version is that the player picking up the last stick is a winner.

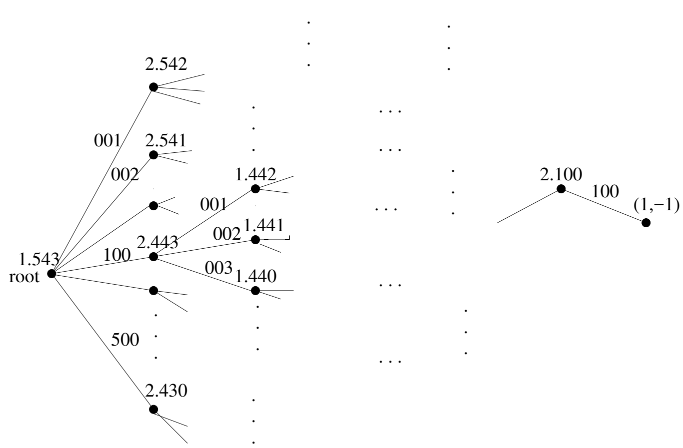


Figure 4.1: Sketch of the extensive form game tree for nim. Many nodes and edges not pictured.

The tree for nim is shown in Figure 4.1. The tree is a rooted directed tree. All edges are directed from left to right; arrowheads are omitted to avoid clutter. The root node is at the left end of the figure. Leaf nodes are the nodes with no outgoing edges, and are labeled with the tuple of payoffs for all players. The only leaf shown in the figure is labeled $(1, -1)$, indicating that player 1 wins: the payoff of player 1 is 1 and the payoff of player 2 is -1. Play of the game generates a path through the tree, beginning at the root and ending at a leaf.

Each node, other than the leaves of the tree, are labeled with the state of the game. The state begins with the index of which player is to play next, and the status of the three piles. The player that plays next at a given node is said to control the node. For example, the root node is labeled 1.543, indicating that player 1 is to play next and the three piles have 5, 4, and 3 sticks in them. respectively. The order of the piles is assumed not to matter. Each edge outgoing from a node corresponds to a possible action for the player controlling the node, and the edge points to the next state. For example, the edge outgoing from the root labeled 002 indicates that player 1 removes two sticks from the third pile. The resulting next state is 2.541. Given the game tree, for any node, we could consider a new game that begins in that node instead of beginning at the root node. The game starting in some arbitrary node is called a *subgame* of the original game. It is assumed that all players know the tree, and, for a perfect information game such as nim, each player knows which state the game is in whenever the player needs to select an action.

Nim is an example of a zero-sum perfect information extensive form game. In theory, the value of the game can be computed using backwards induction. The value of the game, such as the reward to player 1, can be computed for the subgame starting from each node, starting from the leaves. This calculation for nim is shown in Figure 4.2. Reason from the end of the game as follows. State 100 is a losing state (i.e. state 1.100

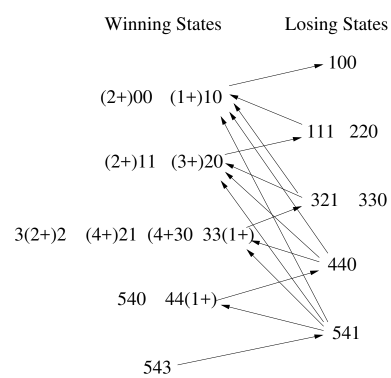


Figure 4.2: Classification of states of nim (misère version)

is a losing state for player 1 and 2.100 is a losing state for player 2) because the player has no other option than to pick up the last remaining stick. That implies that state 200 is a winning state, because a player faced with that state can remove one stick and make the other player face state 100. In fact, any state of the form (2+)00 is a winning state, where "2+" represents a number that is greater than or equal to two. Also, a state of the form (1+)10 is a winning state, where "1+" represents any number in some pile being greater than or equal to one, and the player could remove all sticks from that pile, to again make the other player face state 100.

Next, it must be that 111 and 220 are losing states, because no matter what action a player takes in one of those states, the other player will be left with a winning state. That implies that states of the form (2+)11 and (3+)11 are winning states, and so on. Continuing this process, we find that the initial state, 543, is a winning state. Since player 1 takes the first turn, the value of the game is 1 for player 1 and -1 for player 2.

C. Boutou 1901 found there is a simple characterization of all losing states of nim that works for an arbitrary finite number of piles. Express the sizes of the piles using base two representation, and then add those representations using modulo 2 addition without carries to obtain the *nim sum* of the numbers. For example, the nim sum of 543 is the binary sum of 101 100 011 without carries, or 010. The losing states are those with all nonzero piles having size one and an odd number of ones, such as, 1, 111, 11111, etc. or at least one pile with two or more sticks and nim sum zero. We have described the so-called misère version of the game. The other popular version is the winner is the player to pick up the last stick, in which case the losing states are precisely those with nim sum equal to zero.

Example 4.1 *(Entry deterrence game) Player 1 and player 2 each represent a firm that can sell a product in some market. Player 1 is a potential new entrant to the market, and selects in or out, in other words, to move into the market or stay out. Player 2 has already been in the market for some time, i.e. player 2 is the incumbent. In case player 1 selects in, Player 2 can select accommodate or fight. If player 2 accommodates, the players share the market, and player 2 continues to make profits in the market. If player 2 fights player 1, for example by pricing goods at or below production cost, then the payoffs of both players, especially player 1, are less. The game tree is shown in Figure 4.3. This game has two Nash equilibria in pure strategies:*

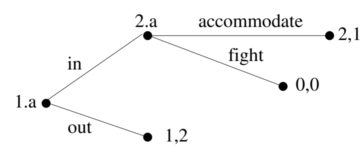


Figure 4.3: Game tree for the entry deterrence game

(out, fight) and (in, accommodate). For the strategy (out, fight), the incumbent is basically declaring that he/she will fight player 1. In other words, player 2 is trying to deter the entrance of player 1. But if player 1 selects in, then, given that decision of player 1, it is against the interests of player 2 to follow through and fight. Since it is costly for the incumbent to fight, given that player 1 selects in, the threat of player 1 to fight might not be considered to be credible. In the terminology given below, only the strategy (in, accommodate) is subgame perfect.

The entry deterrence game illustrates the fact that for a Nash equilibrium of an extensive form perfect information game, the actions selected by some player for some state might not be maximizing the payoff of the player, given that state is reached. Given an extensive form, perfect information game, we consider a pure strategy for a player to be a tuple that gives an action for the player at each of the nodes controlled by the player. A strategy profile in pure strategies, is a tuple of strategies, one for each player.

Definition 4.2 *A strategy profile in pure strategies for an extensive form perfect information game is* subgame perfect *if for each node, the restrictions of the strategies to the nodes of the subgame starting from that node, is a Nash equilibrium of the subgame.*

The subgame perfect Nash equilibria of an extensive form perfect information game can be found by a backwards induction algorithm, as follows. Working backward from the leaves, actions are selected for each node in order to maximize the reward of the player that controls the node. Ties can be broken in an arbitrary way. In this way, a payoff profile is computed at each node for the subgame beginning at that node, corresponding to the actions that have been selected for that node and all other nodes following that node in the game tree. Unless there happens to be a tie, the actions selected at each node will be pure actions and unique. In the special case of two-player zero sum games, all NE are saddle points and thus the payoffs at the root node correspond to the unique value of the game.

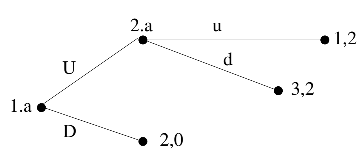
Proposition 4.3 *(Backward induction finds subgame perfect pure strategies) For a perfect information extensive form game, backward induction in pure strategies determines all pure strategy subgame perfect equilibria.*

Proof. Suppose there is a subgame perfect equilibrium. Then by backwards induction proof, it is easy to see that the actions taken could be the output of the backwards induction algorithm.

Conversely, suppose the backward induction algorithm is applied to build up the strategies of all players. We claim that the resulting strategy profile, which is the tuple of strategies, one for each player, produced by the algorithm, is a subgame perfect equilibrium. To prove that, fix any node x in the game tree and show that the strategy profile is a Nash equilibrium for the subgame beginning at x. Fix any player i_o and let $x_1, x_2, \ldots, x_k$ denote nodes in the subtree controlled by player i_o, ordered so that x_i comes before x_j in the game tree if $i \leq j$. Let $s_1, \ldots, s_k$ denote the actions for player i_o in states $x_1, \ldots, x_k$, respectively, as produced by the backwards induction algorithm. We need to show that if $s'_1, \ldots, s'_k$ were alternative actions for player i_o at those nodes then the payoff for player i_o in the subgame beginning at x would be no larger under $s'_1, \ldots, s'_k$ than under $s_1, \ldots, s_k$. The strategies selected by the other players are fixed in this scenario. By backwards induction from the end of the subgame tree, the payoffs for player i_o do not increase as each new action is substituted in and the values are propagated towards the root x of the subgame. ■

Proposition 4.3 generalizes to the case that some nodes in the tree are randomization nodes, i.e. controlled by nature, with an output edge selected randomly by nature for each randomization node. In the presence of randomization nodes, the backwards induction algorithm propagates backward the expected payoff of each player for each node. At each node controlled by a player, the backwards induction algorithm selects an action for that player to steer the state to a node with maximum expected payoff for the player, among the next nodes in the tree.

Remark 4.4 *If any player is faced with a tie when the backwards induction algorithm is run, the choices made by one player to break a tie could influence the payoffs of the other players. Hence, the strategies of all players should be computed by one run of the backwards induction algorithm, from the leaves to the root. It may not be appropriate to run the backwards induction algorithm twice, and then use the strategy found for one player in the first run of the algorithm and the strategy found for another player in the second run of the algorithm. This is illustrated by the following variation of the entry deterrence game:*



The way player 2 breaks the tie at node 2.a influences the optimal choice of player 1.

4.2 Imperfect information extensive form games

4.2.1 Definition of extensive form games with imperfect information, and total recall

To model games with simultaneous actions, such as the prisoners' dilemma problem, and games with hidden information or hidden actions, extensive form games with imperfect information can be considered. Such a game is still specified by a tree, and an outside observer that can see the actions of the players and can also see the outcomes at randomization nodes, can trace a unique path through the tree. But a given player might not perfectly know which state the game is in when the player needs to take an action. Rather, when the player needs to select an action, the player knows that the state of the game is some particular set of states called an *information set*. Thus, the set of all nonleaf nodes in the graph are partitioned into disjoint sets, called information sets, where each set is either an information set controlled by some player, or is controlled by nature. Before giving a definition of extensive form games with imperfect information, we consider an example.

Example 4.5 *(Call my bluff card game) Players 1 and 2 engage in a zero sum game with randomization. First player 1 is dealt a card from a deck of cards. The card is either red, which is good for player 1, or black. Player 1 can see whether the card is red or black, but player 2 can't see the card. The players initially bet one unit of money on the game–in other words, the initial stakes of the game is one unit of money for each player. After seeing the card, the action of player 1 is either to propose to raise the stakes to 2 units of money, or to check, leaving the stakes at one unit of money. If player 1 checks, player 2 has no decision to make. The color of the card is revealed to player 2. If red, player 1 wins one unit. If black, player 1 loses one unit.*

If player 1 proposed to raise the stakes, then player 2 has a decision to make. Player 2 can either meet, meaning player 2 accepts the raise and the stakes of the game are increased to two units of money, or player 2 passes, meaning that player 2 does not accept the raise, but instead drops out of the game. In that case, player 1 wins one unit of money, no matter what the color of the card is; player 2 is not even shown the color of the card. If player 1 raises and player 2 meets, then the color of the card is revealed to player 2. If red, player 1 wins two units of money, if black, player 1 loses two units of money.

The complete game tree is pictured in Figure 4.4. The game starts at node 0 which is a randomization node

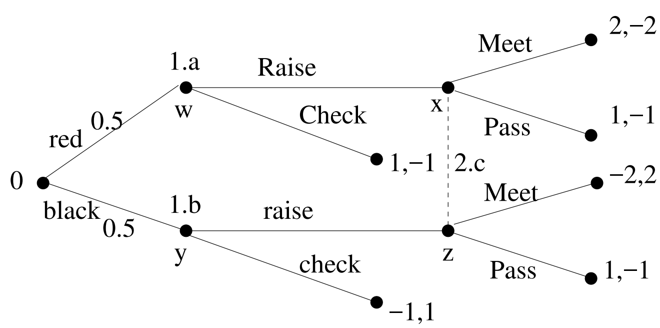


Figure 4.4: Sketch of the extensive form game tree for call my bluff.

controlled by nature. With probability 0.5 nature selects a red card and the game proceeds to node w, and *with probability 0.5 nature selects a black card and the game proceeds to node y. Node w also carries the label $1.a$, where the "1" indicates that the node is controlled by player 1, and the "a" references which decision of player 1 the node corresponds to. Node y is also controlled by player "1." The label $1.b$ on node y is different from the label $1.a$ on node w because player 1 can make different decisions for these two nodes. This reflects the fact player 1 knows the color of the card. If player 1 decides on the action raise from either state 1.a or state 1.b, then the game moves to either state x or z, depending on the color of the card. Player 2 does not know the color of the card; the dashed line connecting states x and z is labeled 2.c, meaning that 2.c, consisting of the set $\{x, z\}$, is an information set for player 2. Player 2 controls the information set and must select an action for the information set, without knowing which state of the set the game is in. Note that the action labels on the edges out of states x and z in the information set $2.c$ are the same, as necessary. The leaf nodes of the game tree are labeled with the corresponding payoff vectors.*

Extensive form games can be used to model single shot games, as illustrated by the following example.

Example 4.6 *(Prisoners' dilemma game in extensive form) The use of information sets to model imperfect information can be used to model one shot games with simultaneous actions as extensive form games. We imagine the players making decisions one at a time, with each player not knowing what decisions were previously made by other players. For example, Figure 4.5 shows the prisoners' dilemma game in extensive form.*

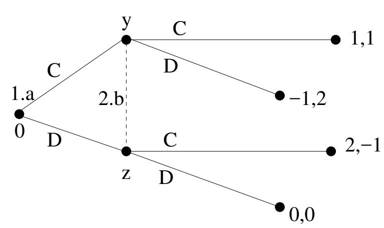


Figure 4.5: Prisoners' dilemma as an extensive game.

An extensive form n-player game with possible imperfect information and randomization is defined as follows. There is a directed tree with a unique root node. The nodes of the trees represent states of the game. Each node and each edge has a unique name or id. Let the players be indexed from 1 to n and use index 0 to denote nature, which is like a player but uses fixed probability distributions rather than decision making. The nodes of the tree are partitioned into information sets, and each information set is controlled by one of the n players or by nature. For any information set, the edges out of any state in the information set are labelled by distinct actions the player could take to move the game along that edge. Furthermore, the same set of action labels is used for all the states of an information set. In addition, if the information set is controlled by nature, a probability is assigned to each action such that the sum of the probabilities over the actions is one.

We will usually restrict attention to extensive form games that have perfect recall, defined as follows.

Definition 4.7 *(Perfect recall) An extensive form game has* perfect recall *if whenever x and x' are nodes in the same information set controlled by some player i, the following is true. Consider the path from the root of the game tree to x, consider the sequence of nodes along that path that are in information sets controlled by player i, and note which action player i takes in each of those information sets along the path. Do the same for the path from the root to x'. Then the sequence of information sets of player i visited and the actions* *taken in those sets is the same for the paths from the root to either x or x'.*

Perfect recall is illustrated in Figure 4.6. Note that x and x' are both in information set $i.k$. The sequence

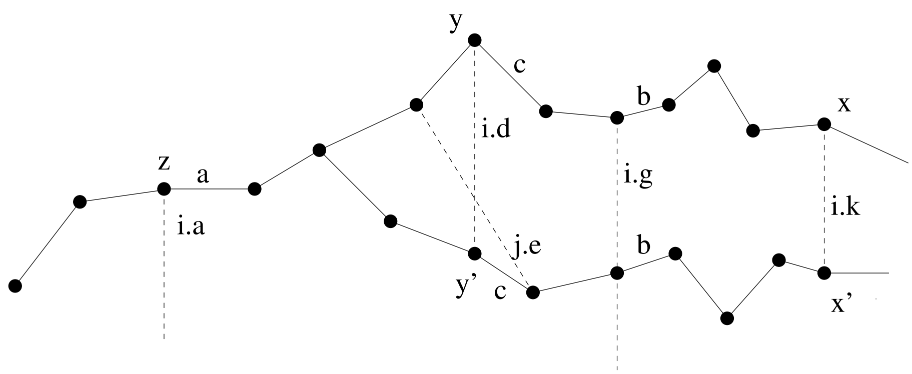


Figure 4.6: Illustration of a game with perfect recall

of information sets of player i and actions taken in those sets, along the path from the root node to either node x or x' is: information set $i.a$, action a, information set $i.d$, action c, information set $i.g$, action b, information set $i.k$. Figure 4.6 also shows the information set $j.e$ of some other player, j. One state in $j.e$ is before a state in $i.d$ and one state in $j.e$ is after a state in $i.d$. Since i and j are different players, this is not a violation of perfect recall.

In order to verify that a game satisfies the perfect recall property, it is sufficient to consider two information sets at a time. Specifically, an extensive form game has perfect recall if, whenever x and x' are nodes in some information set $i.s$ controlled by some player i, and state y is a node along the path from the root to x that is in an information set $i.s'$, also controlled by player i, there must exist a state y' in $i.s'$ that is along the path from the root node to x', such that the action label on the edge out of y along the path to x is the same as the action label on the edge out of y' along the path to x'. For example, in Figure 4.6, x and x' are both in information set $i.k$, and y is along the path from the root to x. So there must be a state y' also in $i.d$ that is along the path to x' and the action labels for the edges out of y and out of y' along the respective paths are the same, namely c. For another example in the same figure, we see that y and y' are both in information set $i.d$ and state z is on the path from the root to y. So there must exist a state z' in the same information set as z such that z' is along the path from the root node to y' and the action labels out of z and z' are the same. In this case, the condition is true with $z' = z$.

Perfect recall implies that the information sets of any single player have a tree order. In other words, the information sets of a player that appear before an information set s of the player are totally ordered, and, furthermore, the actions in each of those information sets that make it possible to reach the next information set (for some possible actions of other players) is unique. This is illustrated in Fig. 4.7, which shows a tree ordering of information sets of player i that is compatible with the game tree partially pictured in Figure 4.6. The actual path taken through the underlying game tree depends on the actions of the other players and nature. The edge in the tree of Fig. 4.7 from information set $i.a$ to information set $i.d$ means that from at least one state in $i.a$, if player 1 plays action a, then it is possible (if actions of other players permit and randomization probabilities are not zero) that the game could reach a state in $i.d$. The tree shown in Fig. 4.7 has a branch point after information set $i.d$. This indicates the possibility for getting from some state in $i.d$ along an edge with action label b to some state in $i.e$, or to some state in $i.f$.

The *normal representation* of an extensive form game is obtained by having each player select a table of

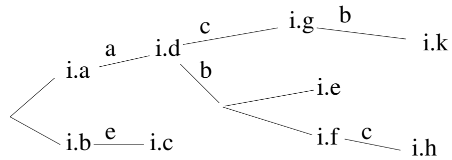


Figure 4.7: Illustration of the tree ordering of information sets of a player implied by the perfect recall property

contingencies–one for each information set under control of the player. For example, for the game corresponding to Fig. 4.7, a pure strategy for player i could look like:

Information set	action
$i.a$	a
$i.b$	e
$i.c$	e
$i.d$	c
$i.e$	c
$i.f$	c
$i.g$	b
$i.h$	e
$i.k$	b

This representation is longer than necessary for some purposes, because, for example, since player i takes action c in information state $i.d$, the game cannot reach any of the information sets $i.e, i.f$, or $i.h$, so the actions given for those states won't be used. A mixed strategy for the normal form representation of the game for a player i is a probability distribution τ_i over the pure strategies of player i. A mixed strategy profile for the game in the normal form representation is a tuple $\tau = (\tau_1, \dots, \tau_n)$ of mixed strategies for each of the players.

Example 4.8 *(Normal form of the call my bluff card game) The game tree for the extensive form version of the call my bluff card game is shown in Fig. 4.4. Player 1 has two information sets. In state 1.a, which is also an information set of cardinality one, which is entered if the card drawn is red, player 1 decides whether to take action R (Raise) or action C (Check). In state 1.b, which is entered if the card drawn is black, player 1 decides whether to take action r (raise) or action c (check). Thus, there are four pure strategies for player 1 in the normal form of the game. Player 2 has one information set that might be reached, namely 2.c, and for the normal form version of the game player 2 must decide what action to take for that information set: either M (Meet) or P (Pass). So the normal form representation of the call my bluff game is given by the following matrix:*

		Player 2	
		M	P
Player 1	Rr	0,0	1,-1
	Rc	.5,-.5	0,0
	Cr	-.5,.5	1,-1
	Cc	0,0	0,0

A mixed strategy for player i is equivalent to a joint probability distribution over the sequence of possible actions for all information sets of player i. Assuming the game has perfect recall, as we do now, allowing arbitrary joint distributions of actions at all information sets is encoding a lot of useless information. For example, referring again to either Fig. 4.6 or 4.7, given that player i has to select an action for information set $i.k$, the player knows that it previously selected actions a, c, and b for information sets $i.a$, $i.d$, and $i.g$, respectively. And it didn't have to select any other actions. It doesn't matter which state of information set $i.k$ the game is in, and the three previous choices by the player were for three information sets, not three underlying states. So given the joint distribution of all actions of the player, we can calculate the conditional probability distribution $\sigma_{i,i.k}$ of the action player i selects for information set $i.k$, given that the player selected actions a, c, and b for information sets $i.a$, $i.d$, and $i.g$, respectively. Such conditional distribution is undefined if the probability of selecting a, c, and b for information sets $i.a$, $i.d$, and $i.g$ is zero, in which case we let $\sigma_{i,i.k}$ be an arbitrary probability distribution over the actions available for information set $i.k$. No matter what strategies the other players use, given that player i has to select an action at state $i.k$, the distribution of that action is the same under the policy τ_i and under $\sigma_{i,i.k}$. We can similarly define $\sigma_{i,s}$ for any information set for player i. Then the tuple of probability distributions for all the information sets of i, σ_i given by $\sigma_i = (\sigma_{i,s} : s \text{ is an information set of } i)$, is called a *behavioral strategy.* And if player i participates in the game by selecting actions independently for visited information sets using σ_i, the distribution of actions made is identical to the distribution of actions made under policy τ_i. We say the policies σ_i and τ_i are behaviorally equivalent.

Any mixed strategy τ_j of any player j has an equivalent behavioral policy σ_j, and the strategy profile σ given by $\sigma = (\sigma_1, \ldots, \sigma_n)$ is behaviorally equivalent for the joint interactions of all the players. In particular, the mean payoffs of all players are the same under τ and σ. We state this result as a theorem.

Theorem 4.9 *(Kuhn's equivalence theorem) For an extensive form game with perfect recall for all players, given a strategy profile τ in mixed strategies for the normal form representation of the game, if one or more of the strategies τ_i is replaced by the behavioral strategies σ_i, the probability distribution of which leaf node in the tree is reached is the same. In particular, the expected payoffs of all players are the same. In other words, payoff equivalence holds.*

It is considerably simpler to describe a probability distribution over actions for each information set than to describe a joint probability over all actions a player would use at all information sets. In view of Kuhn's equivalence theorem, for extensive form games with perfect recall, Nash equilibria represented as profiles of mixed strategies for the normal form game are equivalent to Nash equilibria for profiles of behavioral strategies. In particular, since finite normal form games always have a Nash equilibrium in mixed strategies, we conclude that every extensive form game with perfect recall has a Nash equilibrium in behavioral strategies. We state this as a corollary of Kuhn's equivalence theorem:

Corollary 4.10 *The Nash equilibria of an extensive game with perfect recall in behavioral strategies are profiles consisting of the the behavioral equivalent strategies of the mixed NE of the normal form version of the game.*

Multiagent representation of extensive form game Given an extensive form game, the *multiagent representation* of the game is obtained by replacing each player by a set of players, called agents. There is one agent for each information set of the player. The underlying game tree is kept the same. An information set $i.s$ controlled by player i in the original game becomes the sole information set controlled by the corresponding agent of player i. The payoff vectors at the leaves of the tree for the multiagent representation of the game give the payoffs of each agent, instead of the payoffs of each player, such that the payoff of any agent of player i is the same as the original payoff of the player to which the agent belongs.

A behavioral strategy profile $\sigma = (\sigma_i)_{i\in I}$ for an extensive form game is a selection of behavioral strategies of the players, and the behavioral strategy of each player is a selection of strategies (i.e. probability distributions over available actions) for each of the information sets of the player. Thus, σ has one strategy for each information set in the game. Therefore, σ can be viewed as a strategy profile for the multiagent representation of the game, which itself is a game played by the agents.

Proposition 4.11 *(Nash equilibrium implies multiagent Nash equilibrium) Consider an extensive form game with imperfect information. A Nash equilibrium in behavioral strategies for an extensive form game is also a Nash equilibrium for the multiagent representation of the game.*

The converse of Proposition 4.11 fails, as shown by the following example.

Example 4.12 *(Mutliagent Nash equilibrium does not imply Nash equilibrium for original players) Consider the single player extensive form game of Fig. 4.8. There is a unique Nash equilibrium strategy for player 1*

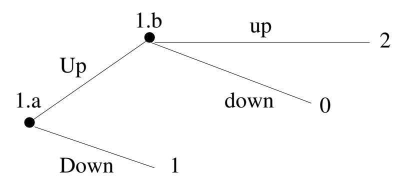


Figure 4.8: Example of a multiagent Nash equilibrium, namely (Down, down), that is not a Nash equilibrium for the original single player game.

in the normal form representation of the game, namely, (Up, up). For the multiagent representation of the game, one agent controls the decision at node 1.a and another agent controls the decision at node 1.b, and the numbers at the leaves give the payoffs of both agents. The strategy pair (Up, up) is indeed a multiagent Nash equilibrium, as guaranteed by Proposition 4.11. However, $(Down, down)$ is also a multiagent Nash equilibrium. Indeed, if the first agent were to change action to Up then the payoff of the first agent would decrease from 1 to 0. If the second agent were to change from down to up, the payoff of the second agent would still be 1. The point of the example is that a single player can change its strategy from (Down, down) to (Up, up) to increase his/her payoff, but no one agent alone could do the same. Of course this example could be extended to make examples involving multiple players.

Example 4.12 suggests that the multiagent representation of the game is not so interesting. However, the mulitagent representation is very useful for a more restrictive definition of equilibrium, as seen in the next section.

4.2.2 Sequential equilibria – generalizing subgame perfection to games with imperfect information

For perfect information games we saw that there is some reason to think that subgame perfect Nash equilibria are more preferable or realistic than other Nash equilibria. In particular, as illustrated in the entry deterrence game, restricting attention to subgame perfect equilibria eliminates some threats that are not credible. The notion of sequential rationality, discussed in this section, is a way to extend the notion of subgame perfect equilibrium to extensive form games with imperfect information.

Example 4.13 *Consider the minor variation of the entry deterrence game of Example 4.1 shown in Fig. 4.9. The edge corresponding to the action "in" in the original game tree is replaced by two edges with*

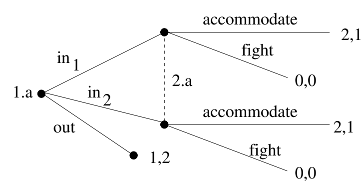


Figure 4.9: The entry deterrence game modified by the addition of a random coin toss

corresponding actions "in$_1$" and "in$_2$," that are equivalent to "in" in the original. Nevertheless, there are now no subgames beginning at a state (i.e. node) of the graph, other than the root node. We thus consider (out, fight) to be a subgame perfect equilibrium. In other words, the concept of subgame perfect based on games starting at individual states of the game don't rule out the not credible threat represented by player 2 using "fight." This example motivates the definition of sequential equilibrium given below – see Remark 4.15.

Information sets are used to model imperfect information in extensive form games. A player must select an action for an information set without knowing which state in the set the game is in. In order to judge whether a player is rationally controlling an information set, it would help to know the probability distribution over the states of the information set the player is assuming. The player might say, "Of course I made a rational decision. Given information set s, I calculated that state x had conditional probability 0.9 and state y had conditional probability 0.1, so I weighted the potential expected payoffs accordingly to select an action for that information set." In order for this statement to be credible, it should be checked that the strategy employed by the player is indeed rational for the stated beliefs, and checked that the stated beliefs are consistent with the game structure and the strategies used by the players and nature. These notions are formalized in what follows.

A *belief vector* $\mu(s)$ for an information set s consists of a probability distribution $(\mu(x|s) : x \in s)$ over the states in the information set. A belief vector μ for the entire game is a vector $\mu = (\mu(s))$ for every information set in the game. A pair (σ, μ) consisting of a behavioral strategy profile σ and a belief vector μ is called an *assessment.*

For an extensive form game with imperfect information, payoffs are determined by which leaf node is reached in the game tree. Let $u_i(\sigma|x)$ denote the expected payoff of player i under strategy profile σ beginning from state x. Note that $u_i(\sigma|x)$ depends only on the decisions made at x and later in the game. For an information set s and belief vector $\mu(s)$, let $u_i(\sigma|s, \mu(s)) = \sum_{x \in s} u_i(\sigma|x)\mu(x|s)$. Also, for a state of the game x and behavioral strategy profile σ, let $p(x|\sigma)$ denote the probability that the game visits state x. Given the behavioral strategy profile σ and an information state s, it is natural to let $\mu(s) = (\mu(x|s) : x \in s)$ be the conditional probability distribution of the state, given that the state is in information set s:

$$\mu(x|s) = \frac{p(x|\sigma)}{\sum_{x' \in s} p(x'|\sigma)} \tag{4.1}$$

The righthand side of (4.1) is well defined if the denominator is greater than zero. In that case, we could consider the pair $(s, \mu(s))$ to be a generalized state of the game. However, the righthand side of (4.1) is not well defined if the denominator is zero, i.e., if the probability of reaching information set s is zero. If we think of large games with perfect information, there might be large numbers of states that have probability zero under a strategy profile, and yet we would like the actions taken by players in those states to be rational in case some changes were made and those states were reached. So by analogy, in games with imperfect information, we would like players to make rational decisions at information sets even if the information sets are off the game path with probability one for the strategy profile under consideration. It would be too arbitrary to allow for arbitrary probability distributions over the set of states in an information set in such cases.

The idea of Kreps and Wilson [12] is to consider distributions that could arise by small perturbations of the policies. A player's strategy in a normal form game is *completely mixed* if it assigns (strictly) positive probability to each of the player's possible actions. Let Σ^o be the set of completely mixed behavioral strategy profiles and let Ψ^o denote the set of assessments of the form (σ, μ) such that $\sigma \in \Sigma^o$ and μ is computed from σ for each information set using (4.1):

$$\Psi^o = \left\{ (\sigma, \mu) : \sigma \in \Sigma^o, \mu(x|s) = \frac{p(x|\sigma)}{\sum_{x' \in s} p(x'|s)} \right\}.$$

Definition 4.14 *(Sequential equilibrium) For an extensive form game with perfect recall:*

(SR) An assessment (σ, μ) is sequentially rational *if for any information set s and any alternative probability distribution σ'_s over the actions at s available to the player i that controls s,*

$$u_i(\sigma'_s, \sigma_{-s}|s, \mu(s)) \leq u_i(\sigma|s, \mu(s)),$$

where σ_{-s} is the set of mixed strategies used by all players at all information sets, except for player i at information set s.

(C) An assessment (σ, μ) is consistent *if there exists $(\sigma^k, \mu^k) \in \Psi^o$ for $k \geq 1$ such that $(\sigma, \mu) = \lim_{k \to \infty}(\sigma^k, \mu^k)$.*

A sequential equilibrium *is an assessment (σ, μ) that is (SR) sequentially rational and (C) consistent.*

Remark 4.15 *(a) Consistency has to do with the beliefs μ being appropriate given σ, and has nothing to do with how rational any player using σ is. For example, σ could correspond to selecting actions that minimize, rather than maximize payoffs. But there is still at least one choice of μ such that (σ, μ) is consistent. The set Ψ^o will often not contain any sequential equilibria because some actions are not rational and should have zero probability. A good thing about Ψ^o is that $(\sigma^k, \mu^k) \in \Psi^o$ means that $\sigma^k \in \Sigma^o$ which means that all nodes of the game tree have positive probability to be reached which means that the corresponding belief vector μ^k is uniquely defined by Bayes rule. All the assessments in Ψ^o are thus consistent in a meaningful sense, and consistency (C) in the definition above means (σ, μ) is in the closure of Ψ^o.*

(b) Let's revisit Example 4.13, which motivates Definition 4.14. In Example 4.13 we had the unsatisfying conclusion that $(out, fight)$ is subgame perfect, even though the example is nearly the same as Example 4.1, for which $(out, fight)$ is not subgame perfect. The problem is that an information set is not a suitable starting node for a subgame. Roughly speaking, by introducing a probability distribution over the states in an information set, we transform an information set into generalized state. Some care has to be taken because the distribution to use is not always uniquely defined. We see that $(out, fight)$ is not a sequential equilibrium for Example 4.13 because for any probability distribution $\mu(2.a)$ over the two states in information set $2.a$, $fight$ is not a rational response for player 2.

(c) *Sequential equilibrium, defined in Definition 4.14, requires (SR) sequential rationality, which involves perturbations of the behavioral strategy of a given player only for one information set at a time. Thus below we will see further use of the multiagent representation of the game.*

(d) *Consider a behavioral strategy profile σ. If an information set s is such that $\sum_{x' \in s} p(x'|s) > 0$, then $\mu(s)$ for any consistent assessment (σ, μ) must satisfy (4.1) because the righthand side of (4.1) is continuous in a neighborhood of σ. Furthermore, if σ is a Nash equilibrium then the behavioral strategy σ_s used for the information state s must be rational for the player using it. In particular, if σ is a fully mixed strategy profile and is a Nash equilibrium, and if μ is defined by (4.1), then (σ, μ) is a sequential equilibrium. Conversely, if (σ, μ) is a sequential equilibrium, σ is a Nash equilibrium.*

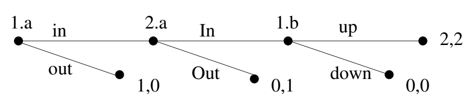
(e) *(Sequential equilibrium generalizes the notion of subgame perfect) A sequential game with perfect information is a special case of a sequential game with imperfect information, such that each information set contains only a single node. For a perfect information game there is only one possible belief vector μ, namely, for each information set s, assigning probability one to the unique node x in s. Moreover, an assessment (σ, μ) is consistent for any behavioral strategy profile σ, and it is a sequential equilibrium if and only if σ is subgame perfect.*

Definition 4.16 *A trembling hand perfect (THP) equilibrium of a finite* normal form *game is a strategy profile $\sigma = (\sigma_i)$ such that there exists a sequence of completely mixed strategy profiles (σ^k) converging to σ such that for each k and each player i, σ_i is a best response to σ^k_{-i} for all k.*

Remark 4.17 *(a) A trembling hand perfect equilibrium is a Nash equilibrium.*

(b) *A trembling hand perfect equilibrium of the normal representation of an extensive form game with perfect information is not necessarily subgame perfect, as shown in the following example. Instead, we consider trembling hand perfect equilibria of the multiagent representation of the game.*

Example 4.18 *Consider the perfect information game shown.*



The strategy profile $((out, down), Out)$, or equivalently, (((0,1),(0,1)), (0,1)), is trembling hand perfect, because it is a vector of best responses for $(((\epsilon, 1-\epsilon), (\epsilon, 1-\epsilon)), (\epsilon, 1-\epsilon))$, for any ϵ with $0 < \epsilon < 1/2$. In other words, $((0,1),(0,1)) \in B_1((\epsilon, 1-\epsilon))$ and $(0,1)) \in B_2(((\epsilon, 1-\epsilon), (\epsilon, 1-\epsilon)))$. However, it is not subgame perfect because "down" is not a rational response for player 1 in the subgame starting at the node with label 1.b. Thus, (σ, μ) is not a sequential equilibrium either. (Recall that perfect information games are special cases of imperfect information games with perfect recall, and there is only one possible belief vector μ. Namely, it assigns probability mass one to the single node within each information set.) However, $((out, down), Out)$, is not a trembling hand perfect equilibrium of the multiagent representation of the game, because "down" is not rational for the second agent of player 1.

Proposition 4.19 *Consider an extensive form game with perfect recall. If σ is a trembling hand perfect equilibrium of the multiagent representation of the game, then there exists a belief system μ such that (σ, μ) is a sequential equilibrium.*

Proof. Suppose σ is a trembling hand perfect equilibrium of the multiagent representation of the game, so there exits a sequence σ^k, $k \geq 1$, of fully mixed behavioral strategy profiles such that $\sigma^k \to \sigma$ and, for each information set s, σ_s is the best response to σ^k_{-s} for the agent controlling state s. Let μ^k denote the belief vector obtained by applying (4.1) with σ replaced by σ^k, so $(\sigma^k, \mu^k) \in \Psi^o$ for all $k \geq 1$. Since the set of all belief vectors is a compact set, there is a subsequence $k_j \to \infty$ as $j \to \infty$ such that $\mu^{k_j} \to \mu$ for some belief vector μ. To avoid messy notation, assume without loss of generality that the entire sequence μ^k is convergent with limit μ, in other words, $\mu^k \to \mu$ as $k \to \infty$. We claim (σ, μ) is a sequential equilibrium. By construction, $(\sigma^k, \mu^k) \to (\sigma, \mu)$, so the assessment (σ, μ) is consistent (C). It remains to show that (σ, μ) is sequentially rational (SR).

By assumption, for any information set s, σ_s is the best response to σ^k_{-s}, which means that for any alternative mixed strategy σ'_s available to the player i controlling information set s,

$$\sum_{x \in S} u_i(\sigma'_s, \sigma^k_{-s}|x)\mu^k(x|s) \leq \sum_{x \in S} u_i(\sigma_s, \sigma^k_{-s}|x)\mu^k(x|s) \tag{4.2}$$

As $k \to \infty$, $u_i(\sigma'_s, \sigma^k_{-s}|x) \to u_i(\sigma'_s, \sigma_{-s}|x)$ and $\mu^k(x|s) \to \mu(x|s)$. Thus, taking the limit $k \to \infty$ on each side of (4.2) yields

$$\sum_{x \in S} u_i(\sigma'_s, \sigma_{-s}|x)\mu(x|s) \leq \sum_{x \in S} u_i(\sigma_s, \sigma_{-s}|x)\mu(x|s),$$

or more concisely, $u_i(\sigma'_s, \sigma_{-s}|x, \mu(s)) \leq u_i(\sigma_s, \sigma_{-s}|x, \mu(s))$. Therefore, the assessment (σ, μ) is (SR) sequentially rational.
■

Proposition 4.20 *(Existence of trembling hand perfect equilibrium (Selton 1975)) A finite normal form game $(I, (A_i : i \in I), (u_i)_{i \in I})$ has at least one trembling hand perfect equilibrium.*

Proof. Let ϵ be so small that $\epsilon < 1/L$, where L is the maximum over all players of the number of actions available to player i. Let T^ϵ denote the set of mixed strategy profiles for the game such that each player plays each action with probability at least ϵ. By the existence theorem for Nash equilibria for convex games, there exists a Nash equilibrium τ^ϵ for the game with strategy profiles T^ϵ, for any $\epsilon > 0$.

Since the set of mixed strategy profiles is compact, there is a sequence $\epsilon^k \to 0$ as $k \to \infty$ such that (τ^{ϵ_k}) converges. Let $\tau = \lim_{k \to \infty} \tau^{\epsilon_k}$. For any player i and action a_i such that $\tau_i(a_i) > 0$ it follows that $\tau_i^{\epsilon_k}(a_i) > 0$ for all k sufficiently large. Since there is a finite set of players, each with a finite set of actions, there is a K large enough that $\tau_i^{\epsilon_k}(a_i) > 0$ for all $k \geq K$ for any player i and action a_i such that $\tau_i(a_i) > 0$. The condition $\tau_i^{\epsilon_k}(a_i) > 0$ implies that a_i is a best response action to $\tau_{-i}^{\epsilon_k}$. So if $k \geq K$, any action for any player i that has positive probability under τ is a best response action for player i against $\tau_{-i}^{\epsilon_k}$. Thus, $(\tau^{\epsilon_{K+k}})_{k \geq 0}$ is a sequence of completely mixed strategy profiles such that τ_i is a best response to $\tau_{-i}^{\epsilon_{K+k}}$ for all $i \in I$ and all $k \geq 0$. Thus, τ is a trembling hand perfect equilibrium of the game. ■

The following is an immediate consequence of Propositions 4.19 and 4.20.

Corollary 4.21 *Every finite extensive form game with perfect recall has at least one sequential equilibrium (σ, μ).*

The following may be of some help in identifying trembling hand equilibria.

Definition 4.22 *(Weakly dominated strategy in normal form game) An action a of a player i in a normal form game is* weakly dominated if there is a mixed strategy α such that $u_i(\alpha, s_{-i}) \geq u_i(a, s_{-i})$ for all s_{-i}

and $u_i(\alpha, s_{-i}) > u_i(a, s_{-i})$ for at least one specific choice of s_{-i}.

Proposition 4.23 *(i) In a trembling hand perfect equilibrium, the strategy of any player assigns zero probability to any weakly dominated action. (ii) The converse is true for two-player games: For two-player games, a Nash equilibrium such that the strategy of neither player is weakly dominated is a trembling hand perfect equilibrium. (See [18, Prop. 248.2]).*

Example 4.24 *(Revisiting the entry deterrence game) The entry deterrence game, described in Example 4.1, is the perfect information extensive form game with the game tree shown in Figure 4.10. It is easy to*

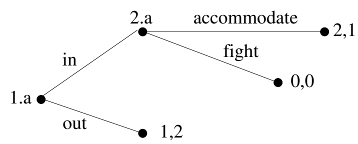


Figure 4.10: Game tree for the entry deterrence game

verify that (in, accommodate) is the unique subgame perfect equilibrium by using Proposition 4.3. But it is instructive to see how Propositions 4.23 and 4.19 can be used to deduce that (in, accommodate) is subgame perfect based on the normal form of the game.

The normal form of the game is given in Table 4.1. This game has two Nash equilibria in pure strategies:

Table 4.1: Normal form of entry deterrence game

		Player 2	
		accommodate	fight
Player 1	in	2,1	0,0
	out	1,2	1,2

(out, fight) and (in, accommodate), as well as some Nash equilibria in mixed strategies. Just looking at the normal form of the game, we see that accommodate *weakly dominates any other strategy for player 2. Thus, by Proposition 4.23, any Nash equilibrium that is THP must have player 2 using accommodate, which implies that (in, accommodate) is (the unique) trembling hand perfect equilibrium for the normal form of the game. Since each player only controls one information set, the original form of the game is equivalent to the multiagent form (for either the sequential or normal representation). Thus, Proposition 4.19 (also see Remark 4.15(c)) implies that (in, accommodate) is a subgame perfect equilibrium.*

4.3 Games with incomplete information

Often when players participate in a game, they lack potentially relevant information about the motivations of the other players. Games of incomplete information (also called Bayesian games) are defined to model this situation in a Bayesian framework.

Definition 4.25 *A (finite) game of incomplete information is given by $G = (I, (S_i)_{i\in I}, (\Theta_i)_{i\in I}, (u_i(s,\theta))_{i\in I}, p(\theta))$ such that*

- *I is the set of players, assumed to be finite*
- *S_i is a finite set of actions available to player i for each $i \in I$, and $S = \times_{i \in I} S_i$ is the set of action profiles, with a typical element $s = (s_i)_{i \in I}$.*
- *Θ_i is a finite set of possible types of player i and $\Theta = \times_{i \in I} \Theta_i$ is the set of type assignments for all players, with a typical element $\theta = (\theta_i)_{i \in I}$.*
- *$u_i(s, \theta)$ is the payoff of player i for given s and θ.*
- *p is a probability mass function (pmf) over Θ (i.e. a joint probability mass function for types). The marginal pmfs, given by $p(\theta_i) = \sum_{\theta_{-i}} p(\theta_i, \theta_{-i})$, are assumed to be strictly positive, so the conditional pmfs, $p(\theta_{-i}|\theta_i)$, are well defined.*

The operational meaning is that at the beginning of the game, nature selects a type assignment θ using pmf p, and each player i learns its own type, θ_i. Players are not told the types of the other players; θ_i is considered to be the private information of player i. However, all players are assumed to know p, so each player i knows $p(\theta_{-i}|\theta_i)$, the conditional probability distribution of the types of the other players. After learning their types, the players each select an action the can depend on their respective types; a pure strategy s_i for a player i is a mapping $s_i : \Theta_i \to S_i$.

A profile of pure strategies $(s_i(\cdot))_{i \in I}$ is a Bayesian (or Bayes-Nash) equilibrium for the game if for each $i \in I$ and each $\theta_i \in \Theta_i$,

$$s_i(\theta_i) \in \arg\max_{a_i' \in S_i} \sum_{\theta_{-i}} p(\theta_{-i}|\theta_i) u_i(a_i', s_{-i}(\theta_{-i}), \theta_i, \theta_{-i}). \tag{4.3}$$

The sum on the righthand side of (4.3) is the expected payoff of player i given the type of the player, θ_i, given player i takes action a_i'. A mixed strategy for player i is a randomized pure strategy. A profile of mixed strategies is a Bayesian (or Bayes-Nash) equilbribum if for each i, the expected payoff of player i is maximized by its randomized strategy, given the randomized strategies used by the other players.

A finite Bayesian game is a special case of an extensive form game with imperfect information. Indeed, given a finite Bayesian game, construct the game with imperfect information as follows. The root node of the game tree would be controlled by nature, which selects $\theta = (\theta_1, \ldots, \theta_n)$, where we assume $I = [n]$. There is one edge of the tree directed out of the root node for each possible assignment of types θ. Then there are n stages of the game tree following the root node, where n is the number of players. The nodes in the first stage are all controlled by player 1. The fact player 1 should know θ_1 can be indicated by partitioning the nodes in stage one into information states controlled by player 1, one for each possible value of θ_1. Each node in stage one has an outgoing edge for each possible action of player 1. In general, the nodes in stage i are partitioned into information sets controlled by player i, one for each possible value of θ_i. When player i makes a decision he/she does not know the types of the other players or the actions taken earlier in the tree by other players.

In turn, there is a normal form version of a Bayesian game. Each player i has a finite set of possible choices of functions $s_i : \Theta_i \to S_i$, and if all players make such a choice the expected payoff of each player is determined by the game. For any finite Bayesian game, a Bayesian equilibrium in mixed strategies exists. This is a consequence of the existence of Nash equilibrium for imperfect information games, or for finite normal form games, which follow from the Kakutani fixed point theorem.

The call my bluff game of Example 4.5 can be viewed as a game of incomplete information by thinking of the color of the card as the type of player 1. Player 2 always has one possible type. An example of an alternative interpretation of the game is the following. We could imagine the players as entering a possible battle, with the strength of the first player being random and known only to the first player. The red type for player 1 means player 1 is strong, and black type means player 1 is weak. The action *raise* for player 1 means player 1 boasts about his/her strength in some way by some provocative action. If player 1 raises, player 2 can decide to either stay in (meet), or surrender (pass).

Example 4.26 *(A Cournot game with incomplete information) Given $c > a > 0$ and $0 < p < 1$, consider the following version of a Cournot game. Suppose the type θ_1 of player 1 is either zero, in which case his production cost is zero, or one, in which case his production cost is c (per unit produced). Player 2 has only one possible type, and has production cost is c. Player 1 knows his type θ_1. It is common knowledge that player 2 believes player 1 is type one with probability p. Both players sell what they produce at price per unit $a - q_1 - q_2$, where q_i is the amount produced by player i. A strategy for player 1 has the form $(q_{1,0}, q_{1,1})$, where q_{1,θ_1} is the amount produced by player one if player 1 is type θ_1. A strategy for player 2 is q_2, the amount produced by player 2.*

The best response functions of the players are given by:

$$\begin{aligned} q_{1,0} &= \arg\max_{q_1} q_1(a - q_2 - q_1) = \frac{(a-q_2)_+}{2} \\ q_{1,1} &= \arg\max_{q_1} q_1(a - c - q_2 - q_1) = \frac{(a-c-q_2)_+}{2} \\ q_2 &= \arg\max_{q_2'} \{(1-p)q_2'(a - c - q_{1,0} - q_2') + pq_2'(a - c - q_{1,1} - q_2')\} \\ &= \arg\max_{q_2'} q_2'(a - c - \overline{q}_1 - q_2') = \frac{(a-c-\overline{q}_1)_+}{2} \end{aligned}$$

where $\overline{q}_1 = (1-p)q_{1,0} + pq_{1,1}$.

Next, we dentify the Bayes-Nash equilibrium of the game, assuming for simplicity that $a \geq 2c$. Substituting the best response functions for player one into the definition of $\overline{q}_1$, yields

$$\overline{q}_1 = \frac{a - pc - q_2}{2}$$

so that we have coupled fixed point equations for $\overline{q}_1$ and q_2. Seeking strictly positive solutions we have

$$q_2 = \frac{(a-c-\overline{q}_1)_+}{2} = \frac{(a - c - \frac{a-pc-q_2}{2})}{2} = \frac{a-(2-p)c+q_2}{4}$$

which can be solved to get

$$q_2 = \frac{a-(2-p)c}{3}, \quad q_{1.0} = \quad \frac{2a+(2-p)c}{6}, \quad q_{1,1} = \frac{2a-(1+p)c}{6},$$

Furthermore, $\overline{q}_1 = \frac{a+(1-2p)c}{3}$.

After finding the equilibrium, it can be shown that the three corresponding payoffs are given by

$$u_2(q^{NE}) = \frac{(a-(2-p)c)^2}{9}$$

and

$$u_1(q^{NE}|\theta_1 = 0) = \frac{(2a+(2-p)c)^2}{36}, \quad u_1(q^{NE}|\theta_1 = 1) = \frac{(2a-(1+p)c)^2}{36}.$$

As expected, $u_2(q^{NE})$ *is increasing in* p*, and both* $u_1(q^{NE}|\theta_1 = 0)$ *and* $u_1(q^{NE}|\theta_1 = 1)$ *are decreasing in* p *with* $u_2(q^{NE}) < u_1(q^{NE}|\theta_1 = 1) < u_1(q^{NE}|\theta_1 = 0)$ *for* $0 < p < 1$*. In the limit* $p = 1$*,* $u_2(q^{NE}) = u_1(q^{NE}|\theta_1 = 1) = \frac{(a-c)^2}{9}$*, as in the original Cournot game with complete information and production cost* c *for both players.*

Chapter 5

Multistage games with observed actions

5.1 Extending backward induction algorithm – one stage deviation condition

A multistage game with observed actions is such that:

- The game progresses through a sequence of stages, possibly infinitely many. Often each stage is also a time period.
- Players select actions simultaneously within a stage.
- Players know the actions of all players in previous stages.

For a finite number of stages these games can be modeled as extensive form games with imperfect information (imperfect because of the simultaneous play within each stage). The following notation will be used. For simplicity, it is assumed that the players use pure strategies, but the propositions hold also for the case of mixed strategies – see Remark 5.8 below.

- $a_i(t)$ is the action of player i at stage t for $1 \leq i \leq n$ and $t \geq 1$, with $a_i(t) \in A_i$, where A_i is a finite set of possible actions for player i at any stage.
- $h_t = (a_i(r) : 1 \leq i \leq n, 1 \leq r \leq t-1)$ is the history at stage t, recording the actions taken strictly before stage t.
- H_t is the set of all tuples of the form $(a_i(r) : 1 \leq i \leq n, 1 \leq r \leq t-1)$ such that $a_i(r) \in A_i$ for all i, r. The set H_t thus includes all possible histories, assuming no restrictions on what actions players can take at each stage.
- A (pure strategy) policy for player i, s_i, maps (stage, history) pairs to action: $s_i(t, \cdot) : H_t \to A_i$ The action $s_i(t, h_t)$ should be defined for all $h_t \in H_t$, even if h_t is not consistent with the values of policy s_i for earlier stages. If policy s_i is used and h_t is the history at stage t, then the action selected by player i is $a_i(t) = s_i(t, h_t)$.

- Payoff functions: $J_i(s_i, s_{-i}) = \sum_{t=1}^{K} u_i(t, s_i(t, h_t), s_{-i}(t, h_t))$, where u_i is the stage payoff function for player i. The horizon K is possibly $+\infty$.

For a given k fixed with $1 \le k \le K$ and history h_k, the subgame for (k, h_k) is the game with payoff functions $J_i^{(k)}(s_i, s_{-i}|h_k) = \sum_{t=k}^{K} u_i(t, s_i(t, h_t), s_{-i}(t, h_t))$. It only depends on s_j's for $t \ge k$. A strategy profile $(s_j : 1 \le j \le n)$ is a *subgame perfect* equilibrium if it is a Nash equilibrium for every subgame (k, h_k).

Example 5.1 *(Trigger strategy for six-stage prisoners' dilemma game) Consider a six-stage game, such that each stage consists of one play of the prisoners' dilemma game. The payoff of player i for the six stage game is the sum of the player's payoffs from the six stages. The payoffs for the prisoners' dilemma game are give by the following table:*

		Player 2	
		C (cooperate)	D
Player 1	C (cooperate)	1,1	-1,2
	D	2,-1	0,0

Strategy D is a dominant strategy for each player for the single shot game.

Consider the following trigger strategy s^T *for player i (the superscript "T" is for "trigger"): Play C in each stage if no player has ever played D earlier. Otherwise, play D. Intuitively, each player is initially being generous, with the idea being that the other player will have incentive to cooperate at each stage because it will improve the options for the other player in future stages. If both players use the trigger strategy, they will both play C in all six stages, so they both receive payoff* 6. *Sounds good. But is* (s^T, s^T) *a Nash equilibrium?*

No matter what happens in the first five stages, for any Nash equilibrium, the best response action for a player in the final stage of the game is to play D. So for any Nash equilibrium, both players play D in stage 6 and both get zero reward in stage 6. Therefore, no matter what happens in the first four stages, for any Nash equilibrium, the best response action for a player in the fifth stage of the game is to play D. So for any Nash equilibrium, both players play D in stage 5 and both get zero reward in stage 5. Continuing with this reasoning for stages 4 down to stage 1, we find that the behavior of the players is unique for any Nash equilibrium for the six stage game, namely, both players play D in every stage, and the payoff vector is $(0, 0)$. *In particular,* (s^T, s^T) *is not a Nash equilibrium. More generally, the behavior of both players is to always play D for any subgame perfect equilibrium.*

Consider the following variation of the example. After each stage, one of the players rolls a fair die to generate a random integer uniformly distributed from 1 to 6, observed by all players. If the number 6 appears, the players stop and the game is over. Else, the players continue to another stage. The total number of stages X is thus random, and it has the geometric distribution with parameter $p = \frac{1}{6}$; $\mathbb{P}\{X = k\} = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}$ for $k \ge 1$ and $\mathbb{E}[X] = 6$. Both players always playing D is again a subgame perfect equilibrium. Are there any other subgame perfect equilibria? How about if both players play the trigger strategy defined above?

We claim the trigger strategy profile (s^T, s^T) is subgame perfect. To prove it, consider an arbitrary subgame (k, h_k). We need to show (s^T, s^T) is a Nash equilibrium for the subgame. The subgame (k, h_k) is reached only if $X \ge k$, so the expected subgame payoffs are computed conditioned on $X \ge k$.

Consider two cases. In the first case, the history h_k indicates at least one of the players selected D at some stage before k, and hence under (s^T, s^T) each player will play D in every stage from stage k until the end. Even if one of the players, i, were to use a different strategy for the subgame, the other player would still play D in every stage of the subgame. Thus, player i would get a strictly smaller payoff whenever he/she played

C instead of D during the subgame. So player i would have no incentive to not use s^T for the subgame.

In the second case, the history h_k indicates that both players selected C in all stages before stage k. Thus, from stage k forward, both players use a trigger strategy starting at stage k just as they would for the entire game starting at stage 1. Also, by the memoryless property of the independent die rolls, the number of stage games that will be played for the subgame (k, h_k) has the same distribution as the number of stages that are played for the original game.

By symmetry we consider only the case that player 2 uses the trigger strategy s^T for the subgame and player 1 uses some other strategy s_1'. Let Y, with $Y \geq k$, denote the first stage player 1 would play D under strategy s_1', if the game doesn't end first. We fix $y \geq k$ and focus on the difference in payoffs on the event $\{Y = y\}$. By averaging over y, by the law of total probability, we can calculate the expected difference of payoffs for Y random. Given $Y = y$, the payoffs for player 1 are the same for s_1 and s_1' up to stage $y - 1$. In particular, if $X \leq y - 1$, the game ends before the policies diverge, and the difference in payoffs is zero. So we can restrict attention to conditioning on the event $\{Y = y, X \geq y\}$ and consider the expected difference in payoffs for stage y until the end of the game. Given $X \geq y$, the conditional distribution of the number of stages remaining starting with state y has the same geometric distribution, with mean 6, as the number of stages in the original game. Thus, the expected payoff of player 1 from stage y onward, given player 1 uses strategy s^T, is 6. If instead player 1 uses s_1' and thus plays D for the first time at stage y, the payoff is 2. Since $2 < 6$, player 1 has a smaller expected payoff under s_1'.

Thus, (s^T, s^T) is subgame perfect as claimed. This reasoning would be true if the probability of continuing at each stage, instead of being 5/6, were anything greater than or equal to 1/2.

Definition 5.2 *A strategy profile s^* satisfies the one-stage-deviation (OSD) condition if for any $i, \bar{k}, \bar{h}_{\bar{k}}$ with $1 \leq \bar{k} \leq K$, if $\widetilde{s}$ agrees with s^* except at $i, \bar{k}, \bar{h}_{\bar{k}}$, then*

$$J_i^{(\bar{k})}(\widetilde{s}_i, s_{-i}^* | \bar{h}_{\bar{k}}) \leq J_i^{(\bar{k})}(s_i^*, s_{-i}^* | \bar{h}_{\bar{k}}). \tag{5.1}$$

Remark 5.3 *If K is finite and deterministic, then s^* satisfies the one stage deviation condition if and only if it can be derived using backwards induction.*

Proposition 5.4 *(One-stage-deviation principle for multistage games with observed actions, K finite) A strategy profile s^* is a subgame perfect equilibrium (SPE) if and only if s^* satisfies the one-stage-deviation condition.*

Proof. (only if) If s^* is an SPE, then by definition, for any $i, \bar{k}, \bar{h}_{\bar{k}}$ with $1 \leq \bar{k} \leq K$, (5.1) is true for any strategy $\widetilde{s}_i$ for player i, including one that differs from s_i^* only at stage $\bar{k}$ for history $\bar{h}_{\bar{k}}$. Thus, s^* satisfies the OSD condition.

(if) Suppose s^* satisfies the OSD. Let i be an arbitrary player and let s_i be an arbitrary strategy for i. To show that s^* is an SPE, it suffices to show that for $1 \leq k \leq K$,

$$J_i^{(k)}(s_i, s_{-i}^* | h_k) \leq J_i^{(k)}(s_i^*, s_{-i}^* | h_k), \text{ for all } h_k. \tag{5.2}$$

We use proof by backwards induction on k. For the base case, note that (5.2) is true for $k = K$, because for each choice of h_K, the only way s_i enters either side of (5.2) is through its value for stage K and history h_K, $s_i(K, h_K)$. Thus, for this case, (5.2) is an instance of the OSD condition, assumed to hold.

For the general induction step, suppose (5.2) is true for $k + 1$, for some k with $1 \leq k \leq K - 1$. Consider an arbitrary choice of h_k, and let h_{k+1} be an extension of h_k using $s_i(k, h_k)$ and $s_{-i}^*(k, h_k)$:

$$h_{k+1} = (h_k, s_i(k, h_k), s_{-i}^*(k, h_k)).$$

Then

$$\begin{aligned} J_i^{(k)}(s_i, s_{-i}^*|h_k) &= u_i(k, s_i(k,h_k), s_{-i}^*(k,h_k)) + J_i^{(k+1)}(s_i, s_{-i}^*|h_{k+1}) \\ &\overset{(a)}{\leq} u_i(k, s_i(k,h_k), s_{-i}^*(k,h_k)) + J_i^{k+1)}(s_i^*, s_{-i}^*|h_{k+1}) \\ &\overset{(b)}{=} J_i^{(k)}(\widetilde{s}_i, s_{-i}^*|h_k) \\ &\overset{(c)}{\leq} J_i^{(k)}(s_i^*, s_{-i}^*|h_k), \end{aligned}$$

where (a) holds by the induction hypothesis, (b) holds for $\widetilde{s}_i$ defined to agree with s_i at (k, h_k) and with s_i^* elsewhere, and (c) holds by the OSD condition. ■

The one step deviation principle can be extended to $K = +\infty$ under the following condition:

Definition 5.5 *The game is continuous at infinity if*

$$\lim_{k\to\infty} \sup_{i,s,\widetilde{s},h_k} \left| J_i^{(k)}(s|h_k) - J_i^{(k)}(\widetilde{s}|h_k) \right| = 0.$$

The game is continuous at infinity, for example, if the stage game payoffs have the form $g_i(s(t), h_t)\delta^{t-1}$ for a bounded function g_i and a discount factor δ with $0 < \delta < 1$.

Proposition 5.6 *(One-stage deviation principle for infinite horizon) Consider the game with infinite horizon, $K = +\infty$.. If the game is continuous at infinity then s^* is a subgame perfect equilibrium (SPE) if and only if the one step deviation (OSD) condition holds.*

Proof. (only if) This part of the proof is the same as for Proposition 5.4 for K finite.

(if) Suppose s^* satisfies the OSD condition. Let $\epsilon > 0$. Fix i and a policy s_i for player i. If suffices to show:

$$J_i^{(k)}(s_i, s_{-i}^*|h_k) \leq J_i^{(k)}(s_i^*, s_{-i}^*|h_k) + \epsilon \text{ for all } h_k \tag{5.3}$$

for all k. By the continuity at infinity condition, there exists $\widetilde{k}$ so large that (5.3) is true for all $k \geq \widetilde{k}$. The backwards induction proof used to prove Proposition 5.4 for K finite can be used to show (5.3) for $0 \leq k \leq \widetilde{k} - 1$ as well, and hence for all k. ■

Example 5.7 *To demonstrate the use of the OSD principle, let's revisit the variation of the trigger strategy for repeated prisoner's dilemma game with random stopping as discussed above. Instead of stopping the game with probability δ after each stage, suppose the game continues for infinitely many stages, but weight the payoffs for stage t by the probability, $(1-\delta)^{t-1}$. that the stage is reached in the original version of the game. Also multiply the payoffs by the constant factor $(1-\delta)$ for convenience, arriving at the stage payoffs: $u_i(t, s(t, h_t)) = (1-\delta)\delta^{t-1} g_i(s_t(t, h_t))$, where g_i is the payoff function for one play of prisoners' dilemma (e.g. $g_1((C,C)) = 1, g_1((C,D)) = -1$, etc.). Let s^T be the trigger strategy of playing C in every stage until at least one player plays D, and then switch to playing D thereafter.*

Let us show that (s^T, s^T) satisfies the OSD condition if $0.5 \leq \delta < 1$, implying (s^T, s^T) is a subgame perfect equilibrium if $0.5 \leq \delta < 1$. Fix $i \in \{1,2\}$ (and let $-i$ denote the other player), fix $\bar{k} \geq 1$ and a history $\bar{h}_{\bar{k}}$. (For example, $i = 1$, $\bar{k} = 4$ and $\bar{h}_4 = ((C,C),(C,D),(D,C))$.) Consider two cases.

Case 1: D appears at least once within $\bar{h}_{\bar{k}}$. *Both players play D at state* $(\bar{k}, \bar{h}_{\bar{k}})$ *under* (s_T, s_T). *So there is only one choice for* $\widetilde{s}$ *in the definition of the OSD condition, namely, player i plays C at state* $(\bar{k}, \bar{h}_{\bar{k}})$ *and* $\widetilde{s}$ *follows* s^T *otherwise. Both players will play D at all stages after stage* $\bar{k}$ *under both* $\widetilde{s}$ *and* s^T *Thus,*

$$J_i^{(\bar{k})}(s_i^*, s_{-i}^*|\bar{h}_{\bar{k}}) - J_i^{(\bar{k})}(\widetilde{s}_i, s_{-i}^*|\bar{h}_{\bar{k}}) = (0 - (-1))(1-\delta)\delta^{\bar{k}-1} > 0.$$

Case 2: D does not appear within $\bar{h}_{\bar{k}}$. *Both players play C at state* $(\bar{k}, \bar{h}_{\bar{k}})$ *under* (s_T, s_T). *So there is only one choice for* $\widetilde{s}$ *in the definition of the OSD condition, namely, player i plays D at state* $(\bar{k}, \bar{h}_{\bar{k}})$ *and* $\widetilde{s}$ *follows* s^T *otherwise. Under* $\widetilde{s}$, *both players will play D after stage k. Thus,*

$$J_i^{(\bar{k})}(s_i^*, s_{-i}^*|\bar{h}_{\bar{k}}) - J_i^{(\bar{k})}(\widetilde{s}_i, s_{-i}^*|\bar{h}_{\bar{k}}) = ((1+\delta+\delta^2+\cdots)-2)(1-\delta)\delta^{\bar{k}-1} = (1-2(1-\delta))\delta^{\bar{k}-1} \geq 0,$$

if $0.5 \leq \delta < 1$.

Thus, the OSD condition is satisfied if $0.5 \leq \delta < 1$, *as claimed.*

Remark 5.8 *Up to this point, this section considers pure strategies, so that, given* i, t, *and* h_t, $s_i(t, h_t)$ *is a particular action in* A_i. *However, Propositions 5.4 and 5.6 readily extend to mixed strategies in behavioral form. Example 4.6 illustrates the fact that single shot (aka normal form) games can be considered to be special cases of extensive form games. By the same idea, multistage games with observed actions and a finite number of stages can also be viewed as examples of extensive form games. Moreover, they are extensive form games with perfect recall. Thus, by Kuhn's equivalence theorem, Theorem 4.9, any mixed strategy* τ_i *for a multistage game, which is a mixture of pure strategies of the form* s_i *considered above, is behaviorally equivalent to a behavioral strategy* σ_i. *The result holds also if* $K = \infty$ *for games that are continuous at infinity. A behavioral strategy for player i,* σ_i, *maps* (*stage, history*) *pairs to* $\mathcal{A}_i$, *where* $\mathcal{A}_i$ *is the set of probability distributions over* A_i. *Thus,* $\sigma_i(t, h_t) \in \mathcal{A}_i$ *for* $1 \leq t \leq K$ *and* $h_t \in H_t$.

The set of histories H_t *is the same whether pure strategies or behavioral strategies are used. If player i uses behavioral strategy* σ_i, *then* $\sigma_i(t, h_t)$ *is the probability distribution for the action* $a_i(t)$ *selected by player i at time t. The other players observe the action* $a_i(t)$ *at time t but they don't use it in selecting the mixed actions* $\sigma_{-i}(t)$ *used at time t. For Nash equilibria we assume the other players know the strategy of player i, and the other players see the same history as player i, so the the other players are assumed to know the probability distribution* $\sigma_i(t, h_t)$ *player i is using at time t at the time they select* $\sigma_{-i}(t, h_t)$.

5.2 Feasibility theorems for repeated games

Let $G = (I, (A_i)_{i\in I}, (g_i)_{i\in I})$ be a finite normal form game and let $0 < \delta < 1$. The repeated game with stage game G and discount factor δ is the normal form game with payoff functions

$$J_i(s) = (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}g_i(s(t, h_t)),$$

where s is a profile of pure strategies, $s = (s_i(t, h_t))_{i\in I, t\geq 1}$. Since $(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1} = 1$, if each player uses a constant strategy $s_i(t, h_t) \equiv a_i$ then $J_i(s) = g_i(a)$ for all i.

In this section we consider behavioral strategies (see Remark 5.8). Let $\mathcal{A}_i$ be the set of probability distributions on A_i, and write α_i for a typical element of $\mathcal{A}_i$ and a_i for a typical element of A_i. A profile of mixed strategies for the stage game has the form $\alpha = (\alpha_i)_{i\in I}$. The interpretation of α is that the players independently select actions, with player i selecting an action in A_i at random with distribution α_i. For a profile of mixed strategies α and a fixed player i, α_{-i} represents a joint distribution over the actions of the other players, such that the actions of the other players are mutually independent. A profile of mixed strategies in behavioral form for the multistage game is a tuple $\sigma = (\sigma_i)_{i\in I}$ such that, for any $k \geq 1$ and history $h_k \in H_k$, $\sigma_i(k, h_k) \in \mathcal{A}_k$. For k and h_k fixed, $(\sigma_i(k, h_k))_{i\in I}$ is a mixed strategy profile for the stage game. In particular, if $(a_i(k))_{i\in I}$ denotes the random actions taken at stage k, the actions are conditionally independent given (k, h_k). The expected reward for player i in the repeated game is written as

$$J_i(\sigma) = (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}g_i(\sigma(t, h_t)),$$

where $g_i(\sigma(t, h_t))$ is the expected reward for player i in the stage game for the profile of mixed strategies $\sigma(t, h_t) = (\sigma_i(t, h_t))_{i\in I}$. In other words,

$$g_i(\sigma(t, h_t)) = \sum_{a\in\times_{i\in I}A_i} g_i(a)\prod_{i\in I}\sigma_i(a_i|t, h_t).$$

Just as for pure strategy profiles, if α is a mixed strategy profile for the stage game, and if σ is the behavioral strategy profile for the repeated game with $\sigma_i \equiv \alpha_i$, then $J(\sigma) = g(\alpha)$.

We have seen that there can be multiple equilibria for the repeated game even if the stage game has a unique Nash equilibrium. For example, it is true if the stage game is prisoners' dilemma. Feasibility theorems address what payoff vectors can be realized in equilibrium for repeated games for δ sufficiently close to one.

Let $\underline{v} = (\underline{v}_i)_{i\in I}$ be the vector of min max values for the respective players in the stage game G, such that each player can use a mixed strategy:

$$\underline{v}_i = \min_{\alpha_{-i}}\max_{\alpha_i}\; g_i(\alpha_i, \alpha_{-i}). \tag{5.4}$$

For any player i, if the other players repeatedly play using $\widehat{\alpha}_{-i} \in \arg\min_{\alpha_{-i}}\max_{\alpha_i}\; g_i(\alpha_i, \alpha_{-i})$, then the other players can ensure $J_i(\sigma) \leq \underline{v}_i$, no matter what strategy σ_i player i uses.

Remark 5.9 *(a) For any player i, α_{-i} can represent any distribution over the actions of the other players with a product form; given α_{-i} the actions of the other players are independent. The set of such product form distributions is not convex, so strong duality can fail. In other words, if there are three or more players, if the order of the min and max in* (5.4) *were reversed, the resulting value could be strictly smaller. For two players the order doesn't matter by the theory of zero sum two-player games, which is what we get when one player is trying to minimize the payoff of the other player.*

(b) Note that $g(\alpha^{NE}) \geq \underline{v}$ (coordinate-wise) for any Nash equilibrium α^{NE} in mixed strategies for the stage game. That is because α_i^{NE} is a best response to α_{-i}^{NE} for player i in the stage game. So $g_i(\alpha^{NE}) = \max_{\alpha_i}\; g_i(\alpha_i, \alpha_{-i}^{NE}) \geq \underline{v}_i$.

Definition 5.10 *A vector v is* individually rational *(IR) if $v_i \geq \underline{v}_i$ for $i \in I$, and strictly IR if $v_i > \underline{v}_i$ for all $i \in I$. A vector v is a* feasible *payoff vector for game G if*

$$v \in \textit{convex hull}\left\{g(\alpha) : \alpha \in \bigtimes_{i\in I} S_i\right\}.$$

Feasible payoff vectors are those that can be achieved for G using jointly random strategies based on publicly available randomness.

Theorem 5.11 *(Nash) If v is feasible and strictly IR then there exists $\overline{\delta} \in (0,1)$ such that for any $\delta \in [\overline{\delta}, 1)$, there exists a Nash equilibrium σ in behavioral strategies so that $v = J(\sigma)$.*

Proof. Suppose v is feasible and strictly IR. By feasibility, there is a probability distribution $(\lambda_a : a \in A)$ over the set of pure strategy profiles a for the stage game G such that $v = E_\lambda[g(a)] = \sum_{a \in A} g(a)\lambda_a$. The probability distribution λ does not necessarily have product form. Suppose that just before actions are to be selected for each stage t, all players learn a variate $a(t)$ generated at random using probability distribution λ[1]. Assume the players use trigger strategies following $(a(t))_{t \geq 1}$. In other words, each player i uses action $a_i(t)$ in each stage t, as long as all other players have done so in all previous stages. If in the first stage such that not all players select actions according to $(a(t))$, there is exactly one player $\widetilde{i}$ not following $(a(t))$, then in all subsequent stages the other players select their respective actions to punish player $\widetilde{i}$ by choosing mixed strategies that force the expected stage payoffs of player $\widetilde{i}$ to be less than or equal to $\underline{v}_{\widetilde{i}}$. Since the mean payoff of each player i in each stage game is v_i, which is strictly greater than $\underline{v}_i$, if δ is sufficiently close to one, no player would have incentive to deviate from the trigger strategy for a one time gain that makes the player lose $v_i - \underline{v}_i$ on average in all subsequent stages. ■

Example 5.12 *(Nash's realization region for the prisoners' dilemma game) For the prisoners' dilemma game of Example 1.1, $\underline{v} = (0,0)$, which happens to be the payoff vector for the unique Nash equilibrium. The set of feasible vectors is the convex hull of the set of possible payoff vectors $\{(1,1), (-1,2), (2,-1), (0,0)\}$ for pure strategy profiles. Nash's realization region, which is the set of feasible, strictly IR vectors, is shown in Fig. 5.1.*

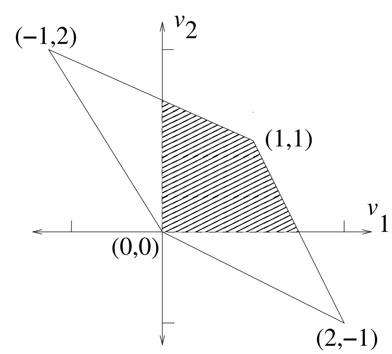


Figure 5.1: Nash realization region (realizable, strictly IR vectors) for prisoners' dilemma game. The region is open from below; it does not include points on the coordinate axes.

A major shortcoming of Theorem 5.11 is that the Nash equilibrium of the repeated game is typically not subgame perfect. In other words, while the players might promise to punish a single player that deviates from the script, if after some number of stages a single player deviates from the script, it could be costly for the other players to follow through and punish the deviating player. The following theorem gives a straightforward way to address this issue.

[1]This assumes the availability of public randomness. It can be shown that if $\overline{\delta}$ is sufficiently close to one, then the same effect as random selection of a can be achieved by using a deterministic time varying schedule with empirical distributions over moderately long time intervals close to λ. When δ is very small, only averages over many stages are relevant and the particular order of events within the stages over a short time period is not critical.

Theorem 5.13 *(Friedman) Let* α^{NE} *be a payoff vector for some Nash equilibrium in mixed strategies of the stage game* G*. If* v *is feasible vector such that* $v_i > v_i^{NE}$ *for all* i*, then there exists* $\overline{\delta} \in (0,1)$ *so for any* $\delta \in [\overline{\delta}, 1)$*, there exists a subgame perfect profile in behavioral strategies* σ *with* $v = J(\sigma)$*.*

Proof. The behavioral strategy profile is constructed as in the proof of Theorem 5.11, but with $\underline{v}$ replaced by v^{NE}. This strategy profile is subgame perfect because for a subgame such that one player first deviated from the script as some time in the past, all players are to be following repeated play of v^{NE}, which is a Nash equilibrium from the subgame. While the Nash equilibrium payoff vector gives punishment for the first player to deviate, it is still provides as large a payoff as available for any single player after the trigger event. ■

Remarkably, under a minor additional assumption, and with a lot of added complexity in the strategies, Nash's original realization theorem, Theorem 5.11, can be implemented in subgame perfect equilibria:

Theorem 5.14 *(Fudenberg and Maskin [9] realization theorem) Suppose the conditions of Nash's realization theorem, Theorem 5.11 hold, and, in addition, suppose the dimension of the set of feasible payoff vectors is equal to the number of players. Then the conclusion of Theorem 5.11 holds for some subgame perfect profile in behavioral strategies,* σ*.*

See [10] for a proof. The idea of the construction of σ, as in the proof of Theorem 5.11, is to incentivize other players to punish the first player to deviate from the the script of a trigger policy. The problem is it may not be in the best interest for the players to follow through and do the punishing, because they may in part be punishing themselves. But then the players acting together can punish players who don't follow through to punish the original players, or reward players that do follow through. The proof is easier to prove under the availability of public randomness, but is still complicated.

Chapter 6

Mechanism design and theory of auctions

Mechanism design and auctions both address the how to allocate resources in exchange for payments. Mechanism design focuses on application specific scenarios in which the agents involved (bidders and seller(s)) may have some information about each other and there may be substantial flexibility for crafting protocols for engagement. The designer of an allocation mechanism is formulating a game, typically in such a way that some game theoretic equilibrium of the game, has desirable properties. For example, a seller might wish to maximize revenue at the Nash equilibrium of an induced game among the bidders. The designer may have in mind specific behaviors of a bidder, given the type or preferences of the bidders.

Auction theory focuses on scenarios with bidders and seller in which the rules of interaction are often selected from among a relatively small set of well known protocols for engagement. The most common types of auctions are second-price auctions (object is sold to highest bidder at price equal to the second highest bid), first-price auctions (object is sold to highest bidder at price equal to the highest bid), and ascending price (a.k.a. English) auctions (object is sold to the highest bidder at price equal to highest bids, and bids are made that increase with time until no bidder wishes to introduce a higher bid).

An important class of mechanisms within the theory of mechanism design are seller mechanisms, which implement the sale of one or more items to one or more bidders. Some authors would consider all such mechanisms to be auctions, but the definition of auctions is often more narrowly interpreted, with auctions being the subclass of seller mechanisms which do not depend on the fine details of the set of bidders. The rules of the threes types of auction mentioned above do not depend on fine details of the bidders, such as the number of bidders or statistical information about how valuable the item is to particular bidders. In contrast, designing a procedure to sell an item to a known set of bidders under specific statistical assumptions about the bidders' preferences in order to maximize the expected revenue (as in [16]) would be considered a problem of mechanism design, which is outside the more narrowly-defined scope of auctions. The narrower definition of auctions was championed by R. Wilson [23].

6.1 Vickrey-Clarke-Groves (VCG) Mechanisms

Let $\mathcal{C}$ represent a set of two or more possible allocations, or social choices, affecting the players indexed by a finite set I. For example,

- $\mathcal{C}$ could be a set of possible locations for a new school in a community.
- Sale of a single item. To model the sale of a single object, such as a license for utilizing a band of wireless spectrum, or a painting, $\mathcal{C}$ could be equal to I, the set of players, with $i \in \mathcal{C}$ denoting that player (or bidder) i gets the object.
- Simultaneous sale of a set of objects. To model the simultaneous sale of a set objects, $\mathcal{O}$, with possibly different objects being distributed to different bidders, $\mathcal{C}$ could be given by

$$\mathcal{C} = \{(A_i)_{i\in I} : A_i \subset \mathcal{O}, A_i \cap A_j = \emptyset\},$$

with a particular $(A_i)_{i\in I} \in \mathcal{C}$ denoting that player i gets the objects in A_i for $i \in I$.

Suppose player i has a valuation function $v_i : \mathcal{C} \to \mathbb{R}$, such that $v_i(c)$ for $c \in \mathcal{C}$ is the value player i places on the allocation c.

Definition 6.1 *An* allocation mechanism M *has the form* $M = ((S_i)_{i\in I}, g, (m_i)_{i\in I})$, *such that*

- S_i *is the set of possible bids of player* i*. Set* $S = \bigtimes_{i\in I} S_i$*, so a bid vector has the form* $s = (s_i)_{i\in I} \in S$.
- $g : S \to \mathcal{C}$; $g(s)$ *is the allocation or social choice determined by the bid vector* s.
- $m_i : S \to \mathbb{R}$; $m_i(s)$ *is the amount of money player* i *has to pay, as determined by the bid vector* s.

A set of valuation functions $(v_i)_{i\in I}$ and an allocation mechanism $M = ((S_i)_{i\in I}, g, (m_i)_{i\in I})$, determines a normal form game $(I, (S_i)_{i\in I}, (u_i)_{i\in I})$, where the action sets S_i are the sets of possible bids for the mechanism, and

$$u_i(s) = v_i(g(s)) - m_i(s). \tag{6.1}$$

In other words, the payoff of player i is how much player i values the allocation, minus how much player i must pay. This form of payoff function is called *quasilinear* because it depends linearly on the payment, m_i.

The basic goal of mechanism design is to devise an allocation mechanism such that when the players engage in the associated game, the equilibrium of the game defined in some sense (for example, the Nash equilibrium or dominant strategy equilibrium) has some desirable property $\mathcal{P}$. Typically desirable properties could be that the revenue (sum of payments) is maximized, or the sum of the valuations of the players is maximized (a version of maximum social welfare). Specifically, for the later example, the (social) welfare for an assignment c is defined by

$$W(c) = \sum_{i\in I} v_i(c)$$

and the set of maximum welfare allocations is $\arg\max_{c\in C} W(c)$. Another consideration is to make it easy for players to decide how much to bid, given their valuation functions. Let $c^* = g(s^{NE})$, such that s^{NE} is the Nash equilibrium the players are anticipated to reach. If the mechanism is such that for any profile of valuation functions $(v_i)_{i\in I}$, the allocation c^* satisfies some property $\mathcal{P}$, then the mechanism is said to *implement* property $\mathcal{P}$ in Nash equilibria.

The Vickrey-Clarke-Groves (VCG) mechanism, also sometimes called the generalized Vickrey mechanism, implements welfare maximization in weakly dominated strategy equilibrium.

Definition 6.2 *(Vickrey-Clarke-Groves (VCG) allocation mechanism) A VCG allocation mechanism for a given allocation problem is $((S_i)_{i\in I}, g, (m_i)_{i\in I})$ such that the space of bids S_i for each player i is the set of possible valuation functions for player i, and allocation and payments are given for a bid vector $(\widehat{v}_i(\cdot))_{i\in I}$ by:*

$$c^* = g(\widehat{v}) \in \arg\max_{c\in\mathcal{C}} \sum_{i\in I} \widehat{v}_i(c), \tag{6.2}$$

$$m_i(\widehat{v}) = -\sum_{j\in I, j\neq i} \widehat{v}_j(c^*) + t_i(\widehat{v}_{-i}) \quad i \in I,$$

where for each $i \in I$, $t_i(\widehat{v}_{-i})$ represents a transfer of money from player i to seller that can depend on the bids of other players but not on the bid of player i.

In words, (6.2) means the VCG allocation rule is to select an allocation that maximizes the sum of the valuation functions reported in the bids of the players.

Proposition 6.3 *(Truthfulness and welfare maximization for VCG mechanisms) Bidding v_i (i.e. letting the bid $\widehat{v}_i$ be equal to the true valuation function v_i, or to v_i plus a constant[1]) is a weakly dominant strategy for player i for each $i \in I$. If all players bid truthfully, the allocation c^* maximizes the welfare.*

Remark 6.4 *(a) $m_i(\widehat{v})$ depends directly on the bids $\widehat{v}_{-i}$ of other players, but it depends on $\widehat{v}_i$ only through the fact the bid $\widehat{v}_i$ influences the selection of c^**

(b) Above, $\widehat{v}_{-i}$ represents the vector of bid functions, $\widehat{v}_{-i} = (\widehat{v}_j : j \in I\backslash\{i\})$, not just the values of the functions at some point. So $\widehat{v}_{-i}$ does not depend on the bid, $\widehat{v}_i$, of player i.

(c) A common choice for transfer payments is

$$t_i(\widehat{v}_{-i}) = \max_{c\in\mathcal{C}} \sum_{j\in I\backslash\{i\}} \widehat{v}_j(c).$$

In other words, t_i is taken to be the maximum social welfare of the set of other players, based on their reported bids. For this choice of t_i, the payment function for player i becomes

$$m_i(\widehat{v}) = \left(\max_{c\in\mathcal{C}} \sum_{j\in I\backslash\{i\}} \widehat{v}_j(c)\right) - \sum_{j\in I\backslash\{i\}} \widehat{v}_j(c^*). \tag{6.3}$$

This gives $m_i(\widehat{v}) \geq 0$, and the payment is the loss in total welfare experienced by the set of other players due to the participation of player i in the game.

Proof. The payoff of player i is given by

$$\begin{aligned} u_i(\widehat{v}) &= v_i(c^*) - m_i(\widehat{v}) \\ &= \left(v_i(c^*) + \sum_{j\in I\backslash\{i\}} \widehat{v}_j(c^*)\right) - t_i(\widehat{v}_{-i}) \end{aligned} \tag{6.4}$$

[1]The VCG selection and payoff rules are unchanged if the bid $\widehat{v}_i$ of a player i is changed by an additive constant, so consider $\widehat{v}_i$ and $\widehat{v}'_i$ to be equivalent if there is a constant C so that $\widehat{v}'_i \equiv \widehat{v}_i + C$.

The last term on the righthand side of (6.4) does not depend on the bid $\widehat{v}_i$ of player i. Thus, player i would like to submit a bid $\widehat{v}_i$ to maximize the sum within the large parentheses on the righthand side of (6.4). That sum is like the social welfare, except it is computed using the bids $\widehat{v}_j$ of the other players, which are not necessarily equal to their true valuation functions. Since the choice of c^* is given by the VCG allocation rule (6.2), if player i were to report truthfully, in other words, if $\widehat{v}_i = v_i$, then the allocation rule would be selecting c^* to maximize the sum within the large parentheses on the righthand side of (6.4). In other words, no matter what bids are submitted by the other players, truthful reporting by player i will make the optimization problem solved by the allocation rule the same as the problem of maximizing the payoff of player i. Thus, player i would have no incentive to not report truthfully, no matter what bids are submitted by the other players. The proof that any nontruthful bid by player i can yield a strictly smaller payoff to player i for some choice $\widehat{v}_{-i}$ is relegated to Appendix 6.1.1. The last statement of the proposition follows from the form of the VCG allocation rule. ■

Example 6.5 *(Specialization of VCG to Vickrey second price auction) To model the sale of a single item take $\mathcal{C} = I$. If bidder i has value v_i for the object then $v_i(c) = v_i \mathbf{1}_{\{c=i\}}$. Each bidder i reports a value $\widehat{v}_i$, by reporting the function $c \mapsto \widehat{v}_i \mathbf{1}_{\{c=i\}}$. The VCG allocation rule* (6.2) *becomes*

$$\begin{aligned} c^* = g(\widehat{v}) &= \arg\max_{c \in I} \sum_{i \in c} \widehat{v}_i \mathbf{1}_{\{c=i\}} \\ &= \arg\max_{c \in I} \widehat{v}_c \end{aligned}$$

and the payment, for the rule (6.3) *described in Remark 6.4, is* $\max_{c \in I \setminus \{c^*\}} \widehat{v}_g$. *In other words, the object is sold to a highest bidder and the price is the highest bid of the remaining bidders.*

Example 6.6 *(VCG for simultaneous sale of objects) Let $\mathcal{O}$ denote a set of objects to be distributed among a finite set of bidders indexed by I. The set of possible allocations is $\mathcal{C} = \{(A_i)_{i \in I} : A_i \subset \mathcal{O}, A_i \cap A_j = \emptyset\}$.*[2] *The valuation function of a bidder i, v_i, is a function of the set of objects, A_i, assigned to bidder i. Write $v_i(A_i)$ to denote the value of set of objects A_i for $A_i \subset \mathcal{O}$, and let $A = (A_i)_{i \in I}$ denote an element of $\mathcal{C}$.*

The VCG allocation rule is given by

$$A^* = g((\widehat{v}_i)) = \arg\max_{A \in \mathcal{C}} \sum_{i \in I} \widehat{v}_i(A_i),$$

(unfortunately computing the allocation is an NP hard problem) and the payment rule (6.3) *described in Remark 6.4, is:*

$$m_i((\widehat{v}_j)) = \left(\max_{A \in \mathcal{C}} \sum_{j \in I \setminus \{i\}} \widehat{v}_j(A_j) \right) - \sum_{j \in I \setminus \{i\}} \widehat{v}_j(A_j^*). \tag{6.5}$$

The payment m_i for a particular bidder i depends on $(\widehat{v}_j)_{j \neq i}$ and A^*, but it depends on the bid $\widehat{v}_i$ only through the fact it allowed A^* to be selected by the mechanism.

A useful interpretation of the payment rule is based on rearranging it to get

$$m_i + \sum_{j \in I \setminus \{i\}} \widehat{v}_j(A_j^*) = \max_{A \in \mathcal{C}} \sum_{j \in I \setminus \{i\}} \widehat{v}_j(A_j). \tag{6.6}$$

---

[2] This definition of $\mathcal{C}$ doesn't require that all objects are sold by the seller.

The following modification of $\widehat{v}_i$ *is a* single-minded bid *that calls for assignment of either* A_i^* *or* $\emptyset$*:*

$$\widehat{v}_i'(A_i) = \begin{cases} m_i & \text{if } A_i = A_i^* \\ 0 & \text{if } A_i = \emptyset \\ -\infty & \text{else} \end{cases}$$

The lefthand side of (6.6) *is the total welfare for* A^* *if* $\widehat{v}$ *is replaced by* $\widehat{v}'$*. The righthand side of* (6.6) *is the maximum total welfare achievable if* $\widehat{v}$ *is replaced by* $\widehat{v}'$ *subject to* $A_i = \emptyset$*. Moreover,* (6.6) *implies that* m_i *is the minimum value that can be used in the definition of the single-minded bid, such that* A^* *maximizes total welfare if* $\widehat{v}_i$ *is replaced by the single-minded bid. Thus, in a sense, the payment function* m_i *is a* minimum to win *rule.*

Example 6.7 *(Example of Example 6.6) Suppose a VCG mechanism in the previous example is applied to sell the objects in* $\mathcal{O} = \{a, b\}$ *to three bidders. A bidder can buy none, one, or both of the objects. Suppose the bids are (for brevity we write* v *instead of* $\widehat{v}$*):*

$$v_1(\emptyset) = 0, \ v_1\{a\} = 10, \ v_1\{b\} = 3, \ v_1\{a, b\} = 13$$
$$v_2(\emptyset) = 0, \ v_2\{a\} = 2, \ \ v_2\{b\} = 8, \ v_2\{a, b\} = 10$$
$$v_3(\emptyset) = 0, \ v_3\{a\} = 3, \ \ v_3\{b\} = 2, \ v_3\{a, b\} = 14$$

Let's determine the assignment of objects to bidders and the payments of the bidders using payment rule (6.5)*. If both items are allocated to the same bidder, the maximum welfare is 14. A larger welfare is achieved by allocating the items to different bidders, and the maximum of 18 is achieved when bidder 1 is assigned object* a*, bidder 2 is assigned object* b*, and bidder* 3 *is assigned no object. In other words, the VCG allocation is* $A^* = (\{a\}, \{b\}, \emptyset)$*.*

To find m_1 *using* (6.5)*, note that the maximum welfare allocation for players 2 and 3 alone is 14, achieved by allocating both objects to player 3, whereas the total value to bidders 2 and 3 for allocation* A^* *is 8. Thus,* $m_2 = 14 - 8 = 6$*. Similarly, the maximum welfare allocation for bidders 1 and 3 alone is 14, achieved by allocating both objects to player 3, whereas the total value to bidders 1 and 3 for allocation* A^* *is 10. Thus,* $m_2 = 14 - 10 = 4$*. The payment of player 3,* m_3*, is zero because player* m_3 *is not assigned any object.*

In summary, the VCG allocation profile is $A^* = (\{a\}, \{b\}, \emptyset)$ *and the payment profile is* $(6, 4, 0)$*. The seller collects a total payment of 10. An unsatisfactory aspect of the solution is that bidder 3 could point out (and maybe even file a lawsuit) that he/she bid 14 for the pair* $\{a, b\}$ *and lost, while the mechanism sold* $\{a, b\}$ *for a total payment of only 6+4=10. In other words, if the seller and player 3 were to break away on their own and negotiate a different deal, they could both be better off. In the terminology of coorperative game theory described in Chapter 7, the VCG mechanism does not necessarily generate a payoff profile in the core of the game.*

Example 6.8 *(VCG allocation of a divisible good and one-dimensional bids) Suppose one unit of a divisible good, such as wireless bandwidth or power output, is to be divided among* n *buyers. Suppose the value function of buyer* i *for quantity* x_i *is given by* $v_i(x_i) = w_i \ln x_i$*, where* w_i *is private information to buyer* i *with* $w_i > 0$*. A VCG mechanism is used to determine the allocation* $x = (x_1, \ldots, x_n)$ *(such that* $x_i \geq 0$ *with* $\sum_i x_i = 1$*) and payments* $(m_1, \ldots, m_n)$ *as a function of the bids. To reduce the amount of communication required, each buyer* i *submits a single positive scaler bid* b_i*, which the mechanism interprets as the value function* $\widetilde{v}_i(x_i) = b_i \ln x_i$*. Let's determine the VCG allocation and payment rule, using the payment rule as described in Remark 6.4.*

The VCG allocation maximizes the social welfare based on the reported functions. In other words, it maximizes $\sum_{i=1}^{n} b_i \ln x_i$ over x subject to $\sum_i x_i = 1$. Introducing a Lagrange multiplier for the constraint and solving yields $x_i = \frac{b_i}{\sum_j b_j}$. So the allocation vector is proportional to the bid vector. Equivalently, $x_i = \frac{b_i}{b_i + B_{-i}}$, where $B_{-i} = \sum_{j:j\neq i} b_j$, which better shows how x_i depends on the bid b_i of buyer i. The payment m_i is the maximum social welfare for the other buyers minus their welfare under the allocation with buyer i participating:

$$\begin{aligned} m_i(b) &= \sum_{j:j\neq i} b_j \ln \frac{b_j}{B_{-i}} - \sum_{j:j\neq i} b_j \ln \frac{b_j}{B_{-i}+b_i} \\ &= B_{-i} \ln\left(1 + \frac{b_i}{B_{-i}}\right). \end{aligned}$$

6.1.1 Appendix: Strict suboptimality of nontruthful bidding for VCG

We complete the proof of Proposition 6.3 by showing that for any player i with some true valuation function v_i, for any choice of $\widehat{v}_i$ that is not equivalent v_i, there exists a choice of $\widehat{v}_{-i}$ such that the payoff of player i for bid $\widehat{v}_i$ is smaller than for truthful bidding.

Suppose without loss of generality that $i = 1$ and that $\widehat{v}_1$ is not equivalent to v_1. Hence, there exists $c', c^* \in \mathcal{C}$ and $\epsilon > 0$ such that $\widehat{v}_1(c') - \widehat{v}_1(c^*) = 2\epsilon + v_1(c') - v_1(c^*)$. Define $\widehat{v}_2$ by

$$\widehat{v}_2(c') = v_1(c^*) \quad \widehat{v}_2(c^*) = v_1(c') + \epsilon \quad \widehat{v}_2(c) = -\infty \;\text{ if } c \notin \{c', c^*\}$$

and $\widehat{v}_j \equiv 0$ for any $j \notin \{1, 2\}$. To be definite, consider the VCG mechanism with $t_1(v_{-1}) \equiv 0$. On one hand, if player 1 were to bid truthfully, the VCG mechanism would select c^* because

$$\begin{aligned} v_1(c') + \widehat{v}_2(c') &= v_1(c') + v_1(c^*) \\ v_1(c^*) + \widehat{v}_2(c^*) &= v_1(c^*) + v_1(c') + \epsilon, \end{aligned}$$

and the payoff of player 1 would be $u_1(v_1, \widehat{v}_{-1}) = v_1(c^*) + \widehat{v}_2(c^*) = v_1(c^*) + v_1(c') + \epsilon$. On the other hand, if player 1 were to use $\widehat{v}_1$, the VCG mechanism would select c' because

$$\begin{aligned} \widehat{v}_1(c') + \widehat{v}_2(c') &= \widehat{v}_1(c') + v_1(c^*) = \widehat{v}_1(c^*) + v_1(c') + 2\epsilon \\ \widehat{v}_1(c^*) + \widehat{v}_2(c^*) &= \widehat{v}_1(c^*) + v_1(c') + \epsilon, \end{aligned}$$

and the payoff of player 1 would be $u_1(\widehat{v}_1, \widehat{v}_{-1}) = v_1(c') + \widehat{v}_2(c') = v_1(c') + v_1(c^*)$, which is smaller than for truthful bidding.

6.2 Optimal mechanism design (Myerson (1981))

Suppose a single seller has an object to sell and there are n bidders, indexed by a set I. Suppose bidders have independent private valuations $(X_i)_{i\in I}$ such that X_i is known to take values in some interval of the form $[0, \omega_i]$, with probability density function (pdf) f_i, and cumulative distribution function F_i. Let $\triangle$ denote the space of probability distributions over I. The seller can use a selling mechanism, whch is a triple (B, π, μ) such that

$B = \times_{i\in I} B_i$ is a space of bid vectors, with a typical element written as $b = (b_1, \ldots, b_n)$.

$\pi : B \to \triangle$ is the allocation mechanism, specifying the probability distribution used to select which bidder gets the object.

$\mu : B \to \mathbb{R}^n$ is the payment rule, such that $\mu_i(b)$ is the (expected) payment of bidder i for bid vector b.

A selling mechanism induces a game of *incomplete information* among the bidders. Incomplete information means that bidders do not know the types of the other bidders. The types are represented by the X_i's, which are known by the bidders to be independently distributed, with X_i having known CDF F_i. Thus, as in Section 4.3, we discuss implementation in Bayes-Nash equilibrium.

Following the celebrated paper Myerson [16], we seek a selling mechanism and a specific Bayes-Nash equilibrium for the induced game, such that the expected payoff of the seller is maximized. We assume the seller has the option to not sell the object, and in that case the object has value r to the seller, for a fixed constant r, known to the seller and all bidders. It could be $r = 0$, sometimes called the assumption of *free disposal.* It could also be $r = -\infty$, meaning the seller must always sell the object to avoid an infinite loss. The payoff of the seller is taken to be

$$U_o = \sum_i \mu_i + r\mathbf{1}_{\{\text{object not sold}\}}.$$

There are many choices for the space of bid vectors B. For example, each bidder might be required to submit a bid in $\mathbb{R}$ or a bid in $\mathbb{R}^d$ for some $d \geq 1$ or in some discrete space or in some space of functions. The Bayes-Nash equilibrium strategies of the bidders are functions mapping their private valuations to the bid spaces, and the strategy of each bidder is supposed to be optimal given the strategies of the other bidders. The large space of possibilities for the bid space B makes it hard to imagine how to get started.

However, there is a natural choice for B. Given the value x_i for player i lies in the interval $[0, \omega_i]$, if the set of possible bids for player i were also equal to $[0, \omega_i]$, then perhaps it could be arranged that reporting truthfully by each player is a Bayes-Nash equilibrium, and it is the Bayes-Nash equilibrium that the seller assumes the players will adopt. Let $V = \times_{i\in I}[0, \omega_i]$, so that V denotes the set of possible value vectors for the players. A *direct mechanism* for the set of possible value vectors V is a selling mechanism (B, π, μ) such that $B = V$, or in other words, the space of bid vectors is the space of value vectors. So a direct mechanism for V has the form (V, Q, M) such that $Q : V \to \triangle$ (allocation rule) and $M : V \to \mathbb{R}^n$ (payment rule).

The revelation principle, stated and proved next, is that for the purpose of designing a selling mechanism to maximize the seller's utility (or to achieve some other desirable property $\mathcal{P}$) at a specified Bayes-Nash equilibrium, there is no loss of optimality in restricting attention to direct selling mechanisms and to truthful bidding as the specified Bayes-Nash equilibrium.

Proposition 6.9 *(Revelation Principle) Let I be a set of n bidders with specified distributions of their independent values $(f_i(x_i) : 0 \leq x_i \leq \omega_i)_{i\in I}$, let (B, π, μ) be a selling mechanism for I, and let β be a Bayes-Nash equilibrium for the induced game (so $\beta_i : [0, \omega_i] \to B_i$ for $i \in I$). Also, let $V = \times_{i\in I}[0, \omega_i]$. There is a direct mechanism (V, Q, M) such that reporting truthfully is a Bayes-Nash equilibrium with the same outcomes under (V, Q, M), as the outcomes of β under (B, π, μ). In other words, the two equilibria give the same distribution of winner and same payoffs for every $x \in V$.*

Proof. The assumption $(\beta_i(\cdot))_{i\in I}$ is a Bayes-Nash equilibrium means, by definition, for any player i and any private value x_i of player i,

$$\beta_i(x_i) \in \arg\max_{b_i} \int_{V_{-i}} (\pi_i(b_i, \beta_{-i}(x_{-i}))x_i - \mu_i(b_i, \beta_{-i}(x_{-i})))\, f_{-i}(x_{-i})dx_{-i}. \tag{6.7}$$

Taking b_i to be of the form $b_i = \beta_i(x_i')$, we trivially get the best response value $\beta_i(x_i)$ by taking $x_i' = x_i$. Therefore, (6.7) implies:

$$x_i \in \arg\max_{x_i'} \int_{V_{-i}} (\pi_i(\beta_i(x_i'), \beta_{-i}(x_{-i}))x_i - \mu_i(\beta_i(x_i'), \beta_{-i}(x_{-i})))\, f_{-i}(x_{-i})dx_{-i}. \tag{6.8}$$

The equivalent direct mechanism is obtained by composing the equilibrium mapping β with the allocation and payment rules of (B, π, μ). In other words, let $Q = \pi \circ \beta$ and $M = \mu \circ \beta$. Then (6.8) becomes:

$$x_i \in \arg\max_{x_i'} \int_{V_{-i}} (Q_i(x_i', x_{-i}))x_i - M_i(x_i', x_{-i})))\, f_{-i}(x_{-i})dx_{-i},$$

which shows bidding truthfully is a Bayes-Nash equilibrium for the game induced by the direct mechanism (V, Q, M). The allocation distribution and the payment vector under (B, π, μ) for the Bayes-Nash equilibrium β and any given $x \in V$ are $\pi \circ \beta(x)$ and $\mu \circ \beta(x)$, respectively, which are the same as the allocation distribution and payment vector, $Q(x)$ and $M(x)$, respectively, for the direct mechanism. ■

The proof of Proposition 6.7 is summarized in Figure 6.1. In essence, the direct mechanism builds in the

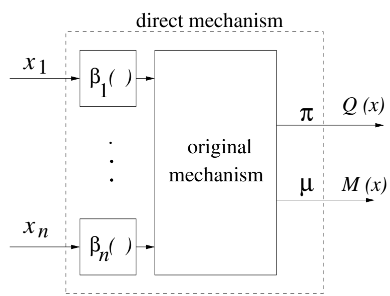


Figure 6.1: Direct mechanism (V, Q, M) is composition of an equilibrium strategy $\beta(\cdot)$ and original mechanism (V, π, μ).

thought process that goes into selecting β. Proposition 6.7 implies we can restrict attention to direct mechanisms and focus on truthful bidding as the Bayes-Nash equilibrium in the induced game. The property of best response being truthful bidding is also known in the mechanism design literature as incentive compatibility:

Definition 6.10 *(Incentive compatibility (IC)) A direct mechanism (V, Q, M) for a space of value vectors V with prior distributions specified by pdfs f_i or CDFs F_i is* incentive compatible *if bidding truthfully is a Bayes-Nash equilibrium.*

The revelation principle still leaves us with huge degrees of freedom for selecting Q and M, but at least it fixes the domains of these functions and it fixes the best response functions of the bidders to be particularly simple, namely, bidding truthfully. To continue on the path to identifying the optimal mechanism, we show next Q and incentive compatibility determine the expected payments, essentially removing M from the optimization.

If bidder i bids z_i he/she gets the object with probability $q_i(z_i)$, where

$$q_i(z_i) = \int_{V_{-i}} Q_i(z_i, x_{-i}) f_{-i}(x_{-i}) dx_{-i}$$

and he/she expects to pay $m_i(z_i)$, where

$$m_i(z_i) = \int_{V_{-i}} M_i(z_i, x_{-i}) f_{-i}(x_{-i}) dx_{-i}.$$

Logically speaking, $q_i(z_i)$ is just a short hand notation for $Q_i(z_i, f_{-i})$, because it is defined by averaging $Q_i(z_i, x_{-i})$ over all values of x_{-i} using the pdf f_{-i}. Similarly, $m_i(z_i)$ is shorthand notation for $M_i(z_i, f_{-i})$. Incentive compatibility can now be stated concisely. The mechanism (V, Q, M) is incentive compatible if and only if, for each $i \in I$ and each $x_i, x_i' \in [0, \omega_i]$,

$$U(x_i) \triangleq q_i(x_i)x_i - m_i(x_i) \geq q_i(x_i')x_i - m_i(x_i'). \tag{6.9}$$

Note that (6.9) represents a linear constraint on the functions q_i and m_i, or on Q_i and M_i, for each pair x_i, x_i'. It turns out these constraints imply q_i is nondecreasing. Moreover, as shown next, the allocation functions (q_i) uniquely determine the expected payment functions (m_i) up to a constant value. In other words, if two IC mechanisms for the same distribution of bidder valuations have the same probability allocation functions q_i, then they have the same payment functions up to additive constants.

Proposition 6.11 *(Revenue equivalence principle) A direct mechanism (V, Q, M) is IC for a set of prior pdfs f if and only if for each $i \in I$, q_i is nondecreasing and m_i is determined (up to an additive constant) by:*

$$m_i(x_i) = m_i(0) + q_i(x_i)x_i - \int_0^{x_i} q_i(t_i)dt_i. \tag{6.10}$$

(If q_i is continuously differentiable, (6.10) *is equivalent to $m_i'(x_i) = q_i'(x_i)x_i$.)*

Proof. (only if) Suppose (V, Q, M) is IC for some choice of pdfs $(f_i)_{i \in I}$. Fix $i \in I$. By the definition of IC, for $x_i, x_i' \in [0, \omega_i]$

$$q_i(x_i)x_i - m_i(x_i) \geq q_i(x_i')x_i - m_i(x_i') \tag{6.11}$$

$$q_i(x_i')x_i' - m_i(x_i') \geq q_i(x_i)x_i' - m_i(x_i) \tag{6.12}$$

Adding the respective sides of (6.11) and (6.12) and rearranging yields

$$(q_i(x_i) - q_i(x_i'))(x_i - x_i') \geq 0,$$

showing q_i is a nondecreasing function.

By incentive compatibility, the function $U_i(x_i) \triangleq q_i(x_i)x_i - m_i(x_i)$ satisfies:

$$U_i(x_i) = \max_{0 \leq y \leq \omega_i} q_i(y)x_i - m_i(y), \tag{6.13}$$

and for x_i fixed, the maximum in (6.13) is achieved at $y = x_i$. Thus, U_i is the maximum of a set of affine functions. By the envelope theorem, Proposition 6.19, it follows U_i is absolutely continuous, $U_i'(x_i) = q_i(x_i)$ for a.e. x_i, and U_i is the integral of its derivative, i.e. $U_i(x_i) = U_i(0) + \int_0^{x_i} q_i(x_i)dx_i$, which is equivalent to (6.10). (Moreover, U_i is a convex function and for any $x_i \in [0, \omega_i,]$, $q_i(x_i)$ is subgradient of U_i at any x_i.)

(if) Conversely, suppose q_i is nondecreasing and m_i satisfies (6.10). The definition $U(x_i) \triangleq q_i(x_i)x_i - m_i(x_i)$ and (6.10) imply $U_i(x_i) = U_i(0) + \int_0^{x_i} q_i(t_i)dt_i$. Together with the assumption q_i is nondecreasing, this implies U_i is convex and for any $x_i' \in [0, \omega_i]$, $q_i(x_i')$ is a subgradient of U_i at x_i. By definition, that means $U_i(x)$ is greater than or equal to the linear function agreeing with U_i at x_i' with slope $q_i(x_i')$. In other words, the inequality in (6.9) holds. ■

Definition 6.12 *(Individually rational) A selling mechanism is* individually rational *(IR) if any bidder i can bid to obtain a nonnegative expected payoff for any value x_i of the bidder.*

Proposition 6.13 *A direct mechanism that is IC is also IR if and only if $m_i(0) \leq 0$ for all bidders i.*

Proof. For a direct mechanism, the expected payoff of bidder i with value x_i and bid x_i' is $q_i(x_i')x_i - m_i(x_i')$. The worst case value of x_i for the bidder is zero, in which case the payoff of the bidder is $-m_i(x_i')$. Therefore, IR holds if and only if there is some choice of x_i' such that $m_i(x_i') \leq 0$. For a direct mechanism with the IC property, m_i is nondecreasing, so the IR property is equivalent to $m_i(0) \leq 0$. ■

Given the allocation rule of a direct mechanism, Proposition 6.11 shows the IC property determines the expected payment functions m_i up to additive constants, and Proposition and 6.13 shows the IR property for an IC payment rule is equivalent to the inequality constraint $m_i(0) \leq 0$. Therefore, to maximize the expected payments, or equivalently, to maximize the expected utility of the seller, subject to the IC and IR constraints and a given allocation rule, it is equivalent to using a payment rule satisfying (6.10) with $m_i(0) = 0$ for all i.

Example 6.14 *Suppose $q_i(x_i) = x_i$ for $0 \leq x_i \leq 1$, so the probability bidder i gets the object is proportional to the bid. What is the maximum expected payment function m_i subject to the IC and IR constraints? Since q_i is differentiable,* (6.10) *becomes $m_i'(x_i) = q_i'(x_i)x_i = x_i$. Solving for m_i with $m_i(0) = 0$, yields $m_i(x_i) = \frac{(x_i)^2}{2}$. This identifies the maximum expected revenue from bidder i for an IC, IR mechanism, given bidder i gets the object with probability x_i for bid $x_i \in [0, 1]$. To check the IR property of this solution, note, for example, if the value of the object to the bidder is 0.6, the expected payoff to the bidder, if the bidder bids x_i, is $x_i(0.6) - \frac{(x_i)^2}{2}$. This payoff is maximized if the bidder bids truthfully, taking $x_i = 0.6$.*

Example 6.15 *If $q_i(x_i) = 1 - e^{-x_i}$ for $x_i \geq 0$, then the expected payment function that maximizes the expected payment (i.e. revenue to the seller) subject to the IC and IR constraints is is given by $m_i'(x_i) = q_i'(x_i)x_i = x_i e^{-x_i}$ and $m_i(0) = 0$, or $m_i(x_i) = 1 - (1 + x_i)e^{-x_i}$.*

Up to this point, given the selection rule Q, we can determine revenue optimal pricing mechanisms subject to the IC and IR constraints. The remaining step towards finding the optimal seller mechanism is to determine Q. We view this as a linear optimization problem, given $(f_i)_{i\in I}$ where f_i is a pdf with support $[0, \omega_i]$ for each $i \in I$. Since the mechanism is constrained to be IC, we maximize the payoff of the seller assuming the bidders bid truthfully. Thus, the optimization problem can be written as:

$$\begin{aligned} &\text{maximize } \ \mathbb{E}\left[\left(\sum_{i\in I} m_i(X_i)\right) + r\mathbf{1}_{\{\text{no bidder gets object}\}}\right] \\ &\text{with respect to } \ (Q, M) \\ &\text{subject to IC and IR} \end{aligned}$$

Since M is determined by Q and the IC and IR constraints, we can write the objective function in a form depending only on the distributions of the values and Q. Taking $m_i(0) = 0$, which is optimal subject to the IC constraints, yields $m_i(x_i) = q_i(x_i)x_i - \int_0^{x_i} q_i(t)dt$.

Therefore,

$$\begin{aligned}\mathbb{E}\left[m_i(X_i)\right] &= \int_0^{\omega_i} q_i(x_i)x_i f_i(x_i)dx_i - \int_0^{\omega_i}\int_0^{x_i} q_i(t)dt f(x_i)dx_i \\ &\overset{(a)}{=} \int_0^{\omega_i} q_i(x_i)x_i f_i(x_i)dx_i - \int_0^{\omega_i}\int_t^{\omega_i} f(x_i)dx_i q_i(t)dt \\ &= \int_0^{\omega_i} q_i(x_i)x_i f_i(x_i)dx_i - \int_0^{\omega_i} (1-F_i(t))q_i(t)dt \\ &= \int_0^{\omega_i} q_i(x_i)x_i f_i(x_i)dx_i - \int_0^{\omega_i} \frac{(1-F_i(x_i))}{f(x_i)} q_i(x_i)f(x_i)dx_i \\ &= \int_0^{\omega_i} \psi_i(x_i)q_i(x_i)f_i(x_i)dx_i \\ &= \mathbb{E}\left[\psi_i(X_i)\mathbf{1}_{\{i \text{ gets object}\}}\right]\end{aligned}$$

where

$$\psi_i(x_i) = x_i - \frac{1-F_i(x_i)}{f_i(x_i)},$$

and $\{i \text{ gets object}\}$ denotes the event player i gets the object. Equality (a) in the above derivation is obtained by changing the order of integration over the region shown in Fig. 6.2.

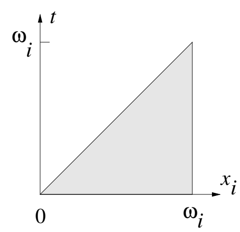


Figure 6.2: A key step in the derivation of the optimal selling mechanism is to change the order of integration, for integration over the region shown.

The function $\psi(x_i)$ is called the *virtual valuation* for bidder i. The total expected revenue is the sum of $\mathbb{E}\left[m_i(X_i)\right]$ over all i, so the mechanism optimization problem reduces to finding the winner selection function Q to solve:

$$\max \mathbb{E}\left[\left(\sum_{i\in I}\psi_i(X_i)\mathbf{1}_{\{i \text{ gets object}\}}\right) + r\mathbf{1}_{\{\text{no bidder gets object}\}}\right] \tag{6.14}$$

with respect to Q

subject to $x_i \mapsto q_i(x_i)$ nondecreasing for each i

Ignore, for a moment, the constraint that q_i be nondecreasing for each i in the optimization problem (6.14). The allocation mechanism selects which bidder gets the object or to have no bidder get the object, after learning $(X_i)_{i\in I}$, so the following selection rule minimizes the quantity inside the expectation with probability one.

MAX VIRTUAL VALUATION selection rule	• Select winner from $\arg\max_i \psi_i(X_i)$ if $\max_i \psi_i(X_i) \geq r$ • Select no winner if $\max_i \psi_i(X_i) < r$

Fortunately, in many cases the constraint on the q_i's is satisfied by this rule, making the rule the optimal solution. Specifically, Myerson defined the mechanism design problem to be *regular* if $\psi(x_i)$ is increasing in x_i for each i. If the problem is regular, then the above selection rule satisfies the constraint that q_i is nondecreasing for each i. If the problem is not regular, the solution is a bit more complicated, with $\psi(x_i)$ being replaced by functions $\widetilde{\psi}(x_i)$ that are nondecreasing. Henceforth we assume the mechanism design problem is regular.

The expected payoff of the seller for the optimal auction is given by

$$\mathbb{E}\left[\max\{\psi_1(X_1), \ldots, \psi_n(X_n), r\}\right]. \tag{6.15}$$

The allocation rule as described above is simple, at least once the functions ψ_i are computed. The above derivation determines the expected payment functions, namely,

$$m_i(x_i) = q_i(x_i)x_i - \int_0^{x_i} q_i(t)dt,$$

but the expected payment $m_i(x_i)$ for a bidder i doesn't depend on whether the player gets the object. Recalling that $M_i(x)$ is the payment of player i given the entire bid vector $x \in V$, any choice of M such that

$$\mathbb{E}\left[M_i(X)|X_i = x_i\right] = m_i(x_i) \tag{6.16}$$

is a valid choice for the revenue optimal mechanism, subject to the IC and IR constraints. Thinking of the second price auction suggests a good choice for M_i. Let $y_i(x_{-i})$ denote the minimum bid for player i such that player i gets the object (assuming the player gets the object in case of a tie):

$$y_i(x_{-i}) = \min\{z_i \in [0, \omega_i] : \psi_i(z_i) \geq r \text{ and } \psi_i(z_i) \geq \psi_j(x_j) \text{ for } j \neq i\}$$

Then consider the payment rule $M_i(x)$ given as follows:

MIN TO WIN payment rule: $M_i(x) = y_i(x_{-i})\mathbf{1}_{\{i \text{ gets object}\}}$.

If the MIN TO WIN payment rule is used, the view of the auction from the standpoint of a bidder i is the same as for Vickrey's second price auction, with $y_i(x_{-i})$ playing the role of the highest bid of the other bidders. In other words, $y_i(x_{-i})$ is a take it or leave it price (that is unknown to the bidder) in the following sense. If x_i is greater than that price, the bidder gets the object for that price. If x_i is less than that price, the bidder does not get the object and pays nothing. Hence, under the MIN TO MIN payment rule, truthful bidding is a weakly dominant strategy. So truthful bidding by all bidders is a Bayes-Nash equilibrium, meaning the mechanism is IC. Also under the MIN TO WIN payment rule, $M_i(0) = 0$, so that $\mathbb{E}\left[M_i(X)|X_i = 0\right] = 0$, which is the maximum possible value of $\mathbb{E}\left[M_i(X)|X_i = 0\right]$ subject to the IR constraint. In summary, the MIN TO WIN rule is IC and IR with maximum possible value (zero) for zero bids. Hence, by the revenue equivalence principle, Proposition 6.11, (6.16) must hold. So the MIN TO WIN rule maximizes the seller's payoff for allocation rule Q subject to the IC and IR constraints.

As a double check, let us verify directly that (6.16) holds for the MIN TO WIN payment rule. Observe

$$q_i(x_i) = \mathbb{P}\left\{i \text{ gets object}|x_i\right\} = \mathbb{P}\left\{y_i(X_{-i}) \leq x_i\right\}.$$

Therefore, q_i is the CDF of the random variable $y_i(X_{-i})$. We use the area rule for expectations of random variables, which for a nonnegative random variable Y bounded above by $y_{\max}$ is $\mathbb{E}\left[Y\right] = \int_0^{y_{\max}} \mathbb{P}\left\{Y > t\right\} dt$, to get

$$\begin{aligned}
\mathbb{E}\left[M_i(X)|X_i = x_i\right] &= \mathbb{E}\left[y_i(X_{-i})\mathbf{1}_{\{y_i(X_{-i})\le x_i\}}\right] \\
&= \int_0^{x_i} \mathbb{P}\left\{y_i(X_{-i})\mathbf{1}_{\{y_i(X_{-i})\le x_i\}} > t\right\} dt \\
&= \int_0^{x_i} \mathbb{P}\left\{t < y_i(X_{-i}) \le x_i\right\} dt \\
&= \int_0^{x_i} (q_i(x_i) - q_i(t))dt \\
&= q_i(x_i)x_i - \int_0^{x_i} q_i(t)dt = m_i(x_i) \qquad (6.17)
\end{aligned}$$

We summarize the above results as a proposition.

Proposition 6.16 *(Myerson(1982)) Given prior pdfs f_i over $[0, \omega_i]$ for the independent private values of bidders indexed by i in I and a known reserve value r of the seller, the seller's expected payoff at Bayes-Nash equilibrium over all IR seller mechanisms is maximized by the direct mechanism (V, Q, M) with Q defined by the MAX VIRTUAL VALUATION selection rule and M defined by the MIN TO WIN payment rule identified above, with the corresponding equilibrium being truthful bidding (so the mechanism is IC).*

Remark 6.17 *The reserve value for bidder i is given by $r_i = \inf\{z_i : \psi(z_i) \ge r\}$. Bidder i can't win if his/her bid is less than r_i. If $x_j < r_j$ for all bidders, the seller does not allocate the object to any bidder.*

Example 6.18 *(Second price auction–alternative verification of IC and IR properties, and constrained optimality) Let us consider some of the implications of this section for second price auctions. As discussed in Examples 1.4 and 6.5, a second price auction is a direct mechanism (V, Q, M) such that $Q(x)$ selects a winner i^* with $i^* \in arg\max_i x_i$ and $M_i(x) = w_i\mathbf{1}_{\{i^*=i\}}$, where $w_i = \max_{j\in I\setminus\{i\}} x_j$.*

As noted in Example 1.4, bidding truthfully is a weakly dominant strategy for any bidder. Truthful bidding is thus also a Bayes-Nash equilibrium for given prior pdfs, $(f_i)_{i\in I}$, on the values of the bidders. In other words, the second price auction is IC. It is also IR, because a bidder can avoid negative payoffs by bidding zero. Moreover, $m_i(0) = 0$ for any bidder, because a player i that bids 0 gets the object with probability zero. While we have thus directly verified the payment rule is IC and $m_i(0) = 0$, let's verify it a second way by appealing to the revenue equivalence principle, Proposition 6.11.

Consider the game from the perspective of some bidder i. Let W_i be a random variable representing the highest bid of the other bidders. Then if bidder i bids x_i, the bidder i gets the object with probability $q_i(x_i) = \mathbb{P}\{W_i \le x_i\}$. In other words, q_i is the CDF of W_i. As in the derivation of (6.17) above, we find the expected payment functions for the second price auction are given by:

$$\begin{aligned}
m_i(x_i) &= \mathbb{E}\left[M_i(X)|X_i = x_i\right] \\
&= \mathbb{E}\left[W_i\mathbf{1}_{\{W_i\le x_i\}}\right] \\
&= x_i q_i(x_i) - \int_0^{x_i} q_i(t)dt.
\end{aligned}$$

Thus Proposition 6.11 implies the second price auction is IC, and we also see $m_i(0) = 0$.

If the mechanism design problem is symmetric, so the pdfs for the values of all bidders are the same, the ψ_i is the same for all bidders. If the problem is also regular, then ψ_i is strictly increasing. Finally, if seller must *sell the object (i.e. $r = -\infty$), then the object goes to the highest bidder in the optimal mechanism. Since this allocation is the same as the allocation of the second price auction, the revenue equivalence principle, Proposition 6.11, implies the second price auction is revenue optimal in the symmetric case when the bidder must sell.*

If the design problem is not symmetric, then the functions ψ_i are not all the same, and choosing i to maximize $\psi(X_i)$ is different from choosing i to maximize X_i. Therefore, the second price auction is not revenue optimal in that case.

6.2.1 Appendix: Envelope theorem

A useful class of functions are the absolutely continuous functions. A function $f : [a, b] \to \mathbb{R}$ is *absolutely continuous* if for any $\epsilon > 0$, there exists $\delta > 0$ such that $\sum_{i=1}^n |f(t_i') - f(t_i)| \leq \epsilon$, for any finite collection of nonoverlapping intervals $\{(t_i, t_i')\}$ in $[a, b]$ such that $\sum_{i=1}^n |t_i' - t_i| \leq \delta$. For example, if f is continuous and piecewise differentiable, then it is absolutely continuous. An absolutely continuous function is differentiable almost everywhere (i.e. at all points of $[a, b]$ except for a set of Lebesgue measure zero). Also, a function $f : [a, b] \to \mathbb{R}$ is an indefinite integral, meaning it can be expressed as $f(x) = f(a) + \int_a^x g(t)dt$ (using Lebesgue integration), if and only if f is absolutely continuous. If f is absolutely continuous, the integrand in the indefinite integral can be any measurable version of $f'(t)$. Here, $f'(t)$ is the derivative of f at t if f is differentiable at t and can be defined arbitrarily otherwise, subject to being a measurable function. (See [20] for proofs.)

Proposition 6.19 (Milgrom and Segal [14]) *Let $\mathcal{F}$ be a set of differentiable and absolutely continuous functions on some interval $[a, b]$. Let $\mathcal{F}^*(t) = \arg\max_{f \in \mathcal{F}} f(t)$, and suppose $\mathcal{F}^*(t) \neq \emptyset$ for almost every $t \in [a, b]$. Also, suppose there is a function $h : [a, b] \to \mathbb{R}$ such that $\int_a^b |h(t)|dt < \infty$ and, for any $f \in \mathcal{F}$, $|f'(t)| \leq h(t)$ almost everywhere. Let $V(t) = \sup\{f(t) : t \in \mathcal{F}\}$ for $t \in [a, b]$. Then V is absolutely continuous, and $V'(t) = f'(t)$ for all $f \in \mathcal{F}^*(t)$, for almost every $t \in [a, b]$.*

Proof. For any interval $(t, t') \subset [0, 1]$, $|V(t) - V(t')| \leq \sup_{f \in \mathcal{F}} |f(t') - f(t)| = \sup_{f \in \mathcal{F}} \int_t^{t'} |f'(s)|ds \leq \int_t^{t'} h(t)dt$, implying that V is absolutely continuous. Therefore, V is differentiable for almost every t. For any $t \in (a, b)$, if V is differentiable at t and $V(t) = f(t)$ for some $f \in \mathcal{F}$, then it must be that $V'(t) = f'(t)$ because $f \leq V$ over $[a, b]$, implying the last part of the proposition. ■

Example 6.20 *Let $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a convex function and let V denote its Legendre-Fenchel transform: $V(t) = \sup_{\theta \in \mathbb{R}} \theta t - g(\theta)$, and let $\theta^*(t) \in \arg\max_\theta \theta t - g(\theta)$. Suppose θ^* is well defined and bounded over some interval $t \in [a, b]$. Then V is absolutely continuous over $[a, b]$ and $V'(t) = \theta^*(t)$ for almost every $t \in [a, b]$. This fact follows from Proposition 6.19 by taking $\mathcal{F}$ to be the set of functions of the form $f(t) = \theta t - g(\theta)$ for $|\theta| \leq \sup\{|\theta^*(t)| : a \leq t \leq b\}$, and $h(t) \equiv \sup\{|\theta^*(t)| : a \leq t \leq b\}$.*

This result can be derived another way under more restrictive assumptions, as follows. Note that $V(t) = \theta^*(t)t - g(\theta^*(t))$. If g and $\theta^*(t)$ were differentiable we could apply the chain rule of differentiation to obtain $V'(t) = \theta^*(t) + (t - g'(\theta^*(t)))\theta^{*\prime}(t) = \theta^*(t)$, where we use the fact $t - g'(\theta^*(t)) = 0$ because of the definition of $\theta^*(t)$.

Proposition 6.19 is limited to functions on $\mathbb{R}$, but otherwise it is rather general. Various envelope theorems for functions with domains of dimension greater than one follow from Proposition 6.19 by considering one-sided directional derivatives of the function in various directions from a given point, or by considering the restriction of such functions to line segments. One such result is the following proposition.

Proposition 6.21 *Let $\mathcal{F}$ be a set of functions $f : D \to \mathbb{R}$, where D is an open subset of $\mathbb{R}^d$. Let $V(t) = \sup\{f(t) : t \in \mathcal{F}\}$ for $t \in D$. Suppose the functions in $\mathcal{F}$ and the function V are continuously differentiable. Let $\mathcal{F}^*(t) = \arg\max_{f \in \mathcal{F}} f(t)$ for $t \in D$ and suppose $\mathcal{F}^*(t) \neq \emptyset$ for all $t \in D$. Then for any $t \in D$, $\nabla V(t) = \nabla f(t)$ for all $f \in \mathcal{F}^*(t)$.*

Chapter 7

Introduction to Cooperative Games

7.1 The core of a cooperative game with transfer payments

Definition 7.1 *A cooperative game with transfer payments consists of* $(I, (v(S))_{S \subset I})$*, where*

- I *is a finite set of players; a subset* S *of* I *is a* coalition *and the set* I *of all players is the* grand coalition*.*
- v *is the* coalitional value function*;* $v(S)$ *is the value or worth of a coalition* S*. It is assumed* $v(S) \in \mathbb{R}$ *for any coalition* S*, and* $v(\emptyset) = 0$*.*

Often in normal form games, a group of players can get better payoffs for themselves by cooperation. For a coalition S, $v(S)$ represents the value the players in the coalition could achieve among themselves, without participation by the players in $I \backslash S$. A central question addressed in the theory of cooperative games is when does there exist a profile of payoffs to the players so that all players have incentive to cooperate within the grand coalition, I, rather than having some subset of players wanting to break away and cooperate instead within a smaller coalition.

Mathematically, the question is whether the *core* of the game is nonempty, where the core is defined as follows. A *payoff profile* is a vector of real numbers, $x = (x_i)_{i \in I}$, such that x_i is the payoff to player i. Let $x(S)$ represent the sum of payoffs for a coalition $S : x(S) = \sum_{i \in S} x_i$.

Definition 7.2 *A payoff profile* x *is* feasible *if* $x(I) = v(I)$*. (In other words, the sum of payoffs is equal to the value of the grand coalition.)*

The core *of the cooperative game* (I, v) *with transferable payoffs is the set of payoff profiles* $(x_i)_{i \in I}$ *such that*

(1) x *is feasible, and*

(2) $x(S) \geq v(S)$ *for all coalitions* S*.*

The feasibility assumption means that the payoffs should be covered by the value of the grand coalition. In other words, there is no subsidy provided to the grand coalition to incentivize the players to cooperate. The condition $x(S) \geq v(S)$ means that if the players in the coalition S decided to break away from the other players, their value $v(S)$, which perhaps they could distribute among themselves, would be less than their sum of payoffs $x(S)$ if they cooperate within the grand coalition.

Next we identify a necessary condition for the core to be nonempty. A *partition* of I is a set of subsets $\{S_1, \ldots, S_K\}$ of I such that $I = \cup_{k=1}^K S_k$ and $S_j \cap S_k = \emptyset$ if $j \neq k$. Each of the sets S_k in a partition is called a block of the partition.

Definition 7.3 *A cooperative game* (I, v) *is* cohesive *if*

$$v(I) \geq \sum_{k=1}^{K} v(S_k) \tag{7.1}$$

for every partition $\{S_1, \ldots, S_K\}$ *of* I.

Equality holds in (7.1) for the partition with only a single block equal to I, because then $K = 1$ and $S_1 = I$. So cohesiveness is equivalent to the condition $v(I) = \max \sum_{k=1}^K v(S_k)$, where the maximum is over all partitions $\{S_1, \ldots, S_K\}$ of I.

If the game is not cohesive, there exists a partition $\{S_1, \ldots, S_K\}$ such that for any feasible payoff profile, $\sum_k x(S_k) = x(I) = v(I) < \sum_k v(S_k)$, implying that $x(S_k) < v(S_k)$ for some block S_k. Therefore, the core of the game is empty if the game is not cohesive. Intuitively, if the game is not cohesive, the players of the grand coalition in aggregate could generate more total value by working within smaller coalitions, so there may be no point to make them cooperate. Of course if some external agency would like to incentivize the players to cooperate in the grand coalition, it could offer to pay a subsidy if all the players agree to cooperate in the grand coalition, effectively increasing the value of $v(I)$, to result in a cohesive game.

In summary, cohesiveness is necessary for the core to be nonempty. Henceforth, we restrict attention to cohesive games. As illustrated in the next example, cohesiveness is not sufficient for the core to be nonempty. However, a theme in this chapter is that for games with many players with enough mixing among them, cohesiveness is typically sufficient for a nonempty core.

Example 7.4 *(I majority game) Consider the coalition game with* $I = \{1, 2, 3\}$ *and* v *defined by* $v\{i\} = 0$[1] *for all* i, $v\{1, 2\} = v\{1, 3\} = v\{2, 3\} = \alpha$ *for some fixed* α *with* $0 \leq \alpha \leq 1$, *and* $v\{1, 2, 3\} = 1$. *The game is cohesive. A vector* (x_1, x_2, x_3) *is in the core if*

$$\begin{aligned} x_i &\geq v\{i\} = 0 \\ x_i + x_j &\geq v\{i, j\} = \alpha \qquad i \neq j \\ x_1 + x_2 + x_3 &= 1 \end{aligned}$$

Thus, if x *is in the core,*

$$1 = x_1 + x_2 + x_3 = \frac{(x_1 + x_2) + (x_1 + x_3) + (x_2 + x_3)}{2} \geq \frac{3\alpha}{2}.$$

So the core is empty if $\alpha > \frac{2}{3}$. *If* $0 \leq \alpha \leq 2/3$, *the core is not empty. For example, it contains* $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. *It also contains* $(1 - \alpha, \alpha, 0)$ *if* $0 \leq \alpha \leq \frac{1}{2}$, *and* $(1 - \alpha, 1 - \alpha, 2\alpha - 1)$ *if* $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$.

Example 7.5 *(production economy) Let* $I = \{c\} \cup W$, *where* c *represents a factory owner, and* W *is a set of* m *workers, for some* $m \geq 1$. *In order for a coalition to have positive worth, it must include the owner and at least one worker, because both the factory and at least one worker are needed for production. Suppose the*

[1]For brevity we write $v\{\cdot\}$ instead of $v(\{\cdot\})$.

worth of a coalition consisting of the owner and k workers is $f(k)$, where $f : [m] \to \mathbb{R}$ is such that $f(0) = 0$, f is monotone increasing, and its increments $f(k+1)-f(k)$ are nonincreasing in k over $0 \le k \le m-1$. In other words, if S_w is a set of workers, $v(S_w) = 0$ and $v(\{c\} \cup S_w) = f(|S_w|)$. The assumption f has nonincreasing increments means that if a worker is added to a coalition that includes the owner, the resulting increase in worth of the coalition is a nonincreasing function of the number of other workers.

Let's find the core of this game. A typical element of the core is a payoff profile of the form $(x_c, x_1, \dots, x_m)$, and the core constraints are (using S_w to denote arbitrary sets of workers):

$$x_c + x_1 + \cdots + x_m = f(m) \tag{7.2}$$

$$x_c + \sum_{i \in S_w} x_i \ge f(k) \text{ if } |S_w| = k \quad 1 \le k \le m \tag{7.3}$$

$$x_i \ge 0 \quad i \in I = \{c\} \cup W \tag{7.4}$$

Constraint (7.3) for $k = m-1$ requires $x_c + \sum_{j \in W \setminus \{i\}} x_j \ge f(m-1)$ for any worker i, which combined with (7.2) implies $x_i \le f(m) - f(m-1)$ for any i. In other words, the payoff to any worker i must be less than or equal to the marginal value of the last worker joining the grand coalition. Thus, the core is contained in the set

$$\{(x_c, x_1, \dots, x_m) : 0 \le x_i \le f(m) - f(m-1) \text{ for } i \in [m] \text{ and } x_c + x_1 + \cdots + x_m = f(m)\} \tag{7.5}$$

However, it is easy to check that any element in the set in (7.5) satisfies the core constraints (7.2)-(7.4), so (7.5) gives the core of the game. The payoff profile in the core that maximizes the payoffs to the workers is $x_c = f(m) - m(f(m) - f(m-1))$ and $x_i = f(m) - f(m-1)$ for $1 \le i \le m$. The payoff profile in the core that minimizes the payoffs to the workers is $x_c = f(m)$ and $x_i = 0$ for $1 \le i \le m$. In other words, the workers get zero payoff.

Definition 7.6 *Let $v = (v(S) : S \subset I)$ with $v(\emptyset) = 0$.*

(a) v is supermodular *if $v(S \cup T) + v(S \cap T) \ge v(S) + v(T)$ for all $S, T \subset I$.*

(b) v is superadditive *if $v(S \cup T) \ge v(S) + v(T)$ for all $S, T \subset I$ such that $S \cap T = \emptyset$.*

Remark 7.7 *(a) Clearly supermodularity implies superadditivity.*

(b) If the coalitional value function v of a cooperative game (I, v) is superadditive, the game is cohesive.

(c) The following is equivalent to supermodularity: $v(A \cup C) - v(A) \ge v(B \cup C) - v(B)$ if $A \cap C = \emptyset$ and $B \subset A$. In other words, the increase in worth for adding C to A is greater than or equal to the increase in worth for adding C to a subset of A. To see the equivalence to the original definition of supermodularity, let $S = A$ and $T = B \cup C$.

The following shows that cooperative games (I, v) with supermodular coalitional value functions v have a nonempty core, and certain elements of the core are easy to identify.

Proposition 7.8 *If (I, v) is a cooperative game such that v is supermodular (some authors call such cooperative games* convex*), then the core is nonempty. Moreover, for any permutation π of I, let $S_k = \{\pi_1, \dots, \pi_k\}$ be the set consisting of the first k elements of π, for $1 \le k \le n$. Then the payoff vector x defined by $x_{\pi_i} = v(S_i) - v(S_{i-1})$ is in the core. (An equivalent expression for x is $x_{\pi_i} = v(S_{i-1} \cup \{x_{\pi_i}\}) - v(S_{i-1})$, so the payoffs are the marginal increases in value as players join to form the grand coalition one at a time in the order of π.)*

Proof. Since the players can be relabeled if necessary, assume without loss of generality that $\pi_i = i$ for $1 \leq i \leq n$. Note that $v(I) = \sum_{i=1}^n v(S_i) - v(S_{i-1}) = \sum_{i=1}^n x_i$, so x is feasible. The remaining requirement, $v(R) \geq x(R)$ for any coalition R, is proved as follows. Suppose $R = \{i_1, \ldots, i_q\}$ such that $i_1 < \ldots < i_q$. By the supermodularity of v and Remark 7.7(c),

$$
\begin{aligned}
v(R) &= \sum_{j=1}^{q} v(\{i_j\} \cup \{i_1, \ldots, i_{j-1}\}) - v(\{i_1, \ldots, i_{j-1}\}) \\
&\leq \sum_{j=1}^{q} v(\{i_j\} \cup S_{i_j-1}) - v(S_{i_j-1}) \\
&= \sum_{j=1}^{q} x_{i_j} = x(R).
\end{aligned}
$$

■

Proposition 7.9 *(Bondareva-Shapley theorem) The core of a cooperative game (I, v) is nonempty if and only if the optimal value for the following linear optimization problem is $v(I)$:*

$$
\begin{aligned}
(D) \quad & \max \sum_{S \subset I} v(S)\lambda_S \\
& \textit{over } \lambda = (\lambda_S)_{S\subset I}, \textit{ subject to} \\
& \qquad \lambda_S \geq 0 \textit{ for all } S \subset I, \\
& \sum_{S: i \in S} \lambda_S \leq 1 \textit{ for all } i \in I
\end{aligned}
$$

(The choice $\lambda_S = \mathbf{1}_{\{S=I\}}$ shows that the optimal value of (D) is greater than or equal to $v(I)$, so the optimal value of the problem being equal to $v(I)$ is the same as the optimal value being less than or equal to $v(I)$.)

Remark 7.10 *If the variables λ in the optimization problem of Proposition 7.9 were restricted to be integer, hence binary, valued, then the optimal value of the problem is $v(I)$ by the cohesiveness assumption. Thus, the problem is a fractionalized version of an integer programming problem connected with cohesiveness. The variable λ_S can be interpreted as the fraction that coalition S is active, and every agent in S needs to devote a fraction λ_S of his/her effort to cooperation within S. The last constraint in problem (D) means the sum of the fractional efforts of any agent i is at most one.*

Proof. The core of a cooperative game (I, v) is nonempty if and only if the optimal value for the following linear optimization problem (P) is $v(I)$:

$$
\begin{aligned}
(P) \quad & \min \sum_{i \in I} x_i \\
& \text{over } x, \text{ subject to} \\
& \sum_{i \in S} x_i \geq v(S) \text{ for all } S \subset I,
\end{aligned}
$$

because the constraints of this problem, along with the condition $\sum_{i\in I} x_i = v(I)$, are exactly the core constraints. In other words, if the value of (P) is $v(I)$, then the core is the set of solutions of (P), and if the core is nonempty, any v in the core is a solution of (P) and the value is $v(I)$.

The proof is completed by showing that the optimization problem (D) in the statement of the proposition is the linear programming dual of problem (P). (Recall that strong duality holds for linear programming problems.) ■

Example 7.11 *Consider the three player majority game of Example 7.4 with* $0 \le \alpha \le 1$. *The vector* $\lambda_S = \frac{1}{2}$ *if* $|S| = 2$ *and* $\lambda_S = 0$ *otherwise, is feasible for the problem (D) in the Bondareva-Shapley theorem. Therefore, the theorem shows that the core being nonempty requires* $\frac{1}{2}v\{1,2\} + \frac{1}{2}v\{2,3\} + \frac{1}{2}v\{3,1\} \le 1$, *or* $3\alpha/2 \le 1$, *or* $\alpha \le \frac{2}{3}$.

Given the definition of cooperative game, it is reasonable to assume that v is superadditive. If v is not superadditive, we could consider the least superadditive majorant, defined as follows.

Definition 7.12 *Given a function* $(v(S) : S \in I)$, *the* least superadditive majorant *of* v *is the minimum function* $\overline{v}$ *such that* $\overline{v}$ *is superadditive and* $\overline{v}(S) \ge v(S)$ *for all* S. *It is given by*

$$\overline{v}(S) = \max \left\{ \sum_k v(S_k) : \{S_1, \ldots, S_K\} \text{ is a partition of } S \right\}.$$

If v is cohesive then v and $\overline{v}$ have the same core. As seen above, if v is not cohesive, then the core of v is empty. In such case it may still be of interest to consider the least superadditive majorant $\overline{v}$ instead, and since $\overline{v}$ is cohesive, it could possibly have a nonempty core.

Given a cooperative game (I, v) and an integer $k \ge 1$, the k-fold replicated game is defined as follows. The set of players is $I' = I \times \{1, \ldots, k\}$, with k players of each type from the original game. For $S' \subset I'$, let

$$w(S') = \begin{cases} v(S) & \text{if for some } S \subset I,\ |S'| = |S|, \text{ and there is one player in } S' \text{ for each type in } S \\ \min_{i \in I} v\{i\} & \text{else (i.e. if } S' \text{ has at least two players of the same type)} \end{cases}$$

and let v' be the least superadditive majorant of w. Note that (I', v') is a cohesive game because v' is superadditive.

Proposition 7.13 *(Koneko & Weiders(1982)) For some integer* K, *the core of the* k-*fold replicated game is nonempty for every* k *that is a positive multiple of* K.

Proof. By the Bondareva-Shapley theorem and the fact that (I', v') is a cohesive game, the core of (I', v') is nonempty if there is a solution to the dual problem for (I', v') such that the λ's are binary valued. The value of the dual problem for (I', v') is K times the value of the dual problem for (I, v). Since the dual problem for (I, v) is a linear optimization problem such that the constraints involve only integer values and integer coefficients, a standard result in the theory of linear programming implies there exists a solution λ^* of the dual problem for (I, v) with rational entries. Thus, for some integer $K \ge 1$, $\lambda^*_S = \frac{k_S}{K}$ for integers $(k_S : S \subset I)$. This implies that there is a solution to the dual problem for (I', v') with binary values of the λ's, which, as noted above, completes the proof. ■

Example 7.14 *Consider the three player majority game of Example 7.4 with* $0 \le \alpha \le 1$. *The two-fold replicated game has players* $I' = \{1, 1', 2, 2', 3, 3'\}$, $v'(I') = 3\alpha = w'(\{1,2\}) + w(\{3,1'\}) + w(\{2',3'\})$. *Note that* $v'(I')$ *is not simply* $2v(I)$. *It is easily checked that* $\left(\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\alpha}{2}\right)$ *is in the core of* (I', v').

Example 7.15 *(The core for simultaneous sale of objects) Examples 6.6 and 6.7 are about use of the VCG mechanism for simultaneous sale of objects. Here we model the sale as a cooperative game and focus on the core. Let $\mathcal{O}$ denote a set of objects to be distributed among a finite set of bidders indexed by I. The set of possible allocations is $\mathcal{C} = \{(A_i)_{i\in I} : A_i \subset O, A_i \cap A_j = \emptyset\}$. Assume the valuation function of a bidder i, v_i, is determined by the set of objects, A_i, assigned to bidder i, and write $v_i(A_i)$ to denote the value of set of objects A_i to bidder i, for $A_i \subset \mathcal{O}$.*

To model this scenario as a cooperative game we take the seller to also be a player, because it doesn't make sense to let any one of the bidders simply break away on his/her own and be able to get whatever bundle of objects he/she would like. To be definite, let us suppose the value of any set of objects to the seller is zero. Thus consider the cooperative game with set of players $I' = \{0\} \cup I = \{0, \ldots, n\}$, where player 0 is the seller. The coalitional value function $v(S)$ is given by:

$$v(S) = \begin{cases} \max_{\textit{partitions } (A_j)_{j\in S\setminus\{0\}}} \sum_{j\in S} v_j(A_j) & \textit{if } 0 \in S \\ 0 & \textit{else} \end{cases} \tag{7.6}$$

A payoff vector x is in the core if and only if $v(S) \le x(S)$ for all S and $x(I) = v(I)$. The core is not empty. For example, $x_0 = v(I')$ and $x_i = -$ for $1 \le i \le n$ (i.e. all value goes to the seller) is in the core.

Let's find the core for the specific case of Example 6.7: $\mathcal{O} = \{a, b\}$ and three bidders with value functions:

$$\begin{aligned} &v_1(\emptyset) = 0,\ v_1\{a\} = 10,\ v_1\{b\} = 3,\ v_1\{a,b\} = 13 \\ &v_2(\emptyset) = 0,\ v_2\{a\} = 2,\ \ v_2\{b\} = 8,\ v_2\{a,b\} = 10 \\ &v_3(\emptyset) = 0,\ v_3\{a\} = 3,\ \ v_3\{b\} = 2,\ v_3\{a,b\} = 14 \end{aligned}$$

The nonzero values of the coalition value function are given by:

$$\begin{aligned} v\{0,1,2,3\} &= 18 \\ v\{0,1,2\} = 18, v\{0,1,3\} &= 14, v\{0,2,3\} = 14 \\ v\{0,1\} = 13, v\{0,2\} &= 10, v\{0,3\} = 14 \end{aligned}$$

The core constraints are thus:

$$\begin{aligned} x_i &\ge 0 \quad \text{for } 0 \le i \le 3 \\ x_0 + x_1 + x_2 + x_3 &= 18 \\ x_0 + x_1 + x_2 &\ge 18 \\ x_0 + x_1 + x_3 &\ge 14 \\ x_0 + x_2 + x_3 &\ge 14 \\ x_0 + x_1 &\ge 13 \\ x_0 + x_2 &\ge 10 \\ x_0 + x_3 &\ge 14 \end{aligned}$$

The first three lines of these constraints imply $x_3 = 0$, and the constraints on the fourth and fifth lines are redundant because of the positivity constraints and the constraint $x_0 + x_3 \ge 14$. Thus, the core constraints *simplify to:*

$$x_0 \geq 14$$
$$x_1, x_2 \geq 0$$
$$x_3 = 0$$
$$x_0 + x_1 + x_2 = 18$$

In words, if x is in the core, the sum of the values must be 18 and at least 14 units of value must be allocated to the seller. The remaining 4 units of value can be shared arbitrarily among the seller and the first two bidders.

Recall that the VCG allocation is to sell item a to bidder 1 and item b to bidder 2 for the payment vector $(6, 4, 0)$. Thus, the payoff vector for the VCG allocation and payment rule is $x^{VCG} = (10, 10-6, 8-4, 0) = (10, 4, 4, 0)$. Thus, the VCG outcome is not in the core. Selecting an allocation in the core eliminates the problem mentioned in Example 6.7, that bidder 3 bid more than the amount charged by the seller, but did not get the object. However, replacing the VCG outcome by an outcome in the core breaks the incentive compatibility property of the VCG mechanism.

7.2 Markets with transferable utilities

An important family of cooperative games are market games. A market is a 4-tuple $M = (I, \ell, (w_i), (f_i))$ consisting of

- I, a set of n agents for some finite n.
- $\ell \geq 1$, the number of goods The goods are assumed to be divisible, such as water, oil, steel, data rate
- $w_i \in \mathbb{R}_+^\ell$, the initial endowment of agent i, for $i \in I$. The entries of w_i are the amounts of each good agent i has initially.
- $f_i : \mathbb{R}_+^\ell \to \mathbb{R}$ is a production/utility/happiness function for agent $i \in I$, assumed to be increasing, continuous, and concave.

We consider markets with transferrable utilities, which means that agents can exchange cash as well as goods. Let m_i be the payment of agent i. A negative value of m_i means that agent i has a net increase in cash. An *outcome* for a given market is a set of final allocations $(z_i)_{i\in I}$, with $z_i \in \mathbb{R}_+^\ell$ for each $i \in I$, and payments $(m_i)_{i\in I}$ satisfying the following constraints:

1. Conservation of goods: $\sum_{i\in I} z_i = \sum_{i\in I} w_i$
2. Budget balanced: $\sum_{i\in I} m_i = 0$

Given an outcome (z, m), the payoff vector $x \in \mathbb{R}^n$ is defined by $x_i = f_i(z_i) - m_i$.

A market induces a cooperative game (I, v) with coalitional value function v defined by

$$v(S) \triangleq \max\left\{\sum_{i\in S} f_i(z_i) : z_i \in \mathbb{R}_+^\ell, \sum_{i\in S} z_i = \sum_{i\in S} w_i\right\}. \tag{7.7}$$

The idea is that the agents can benefit by exchanging goods and money among themselves. The value, $v(S)$, of a coalition $S \subset I$ is the maximum sum of utilities that the agents within the coalition S could achieve by redistributing their initial endowments among themselves. Agent i obtains a profile of goods z_i for each $i \in S$. If $(S_k)_{1\leq k\leq K}$ is a partition of I, then $\sum_{k=1}^{K} v(S_k)$ is the maximum sum of utilities that can be achieved by redistribution of initial endowments, subject to conservation of the amount of each good within each block S_k. The induced cooperative game is cohesive because the maximum sum of utilities over all agents can be achieved by unrestricted redistribution of the initial endowments, conserving only the total amount of each good within the grand coalition I.

Suppose (z, m) is an outcome for M such that the associated payoff vector x is in the core. Then the equality core constraint, namely $\sum_{i\in I} x_i = v(I)$, gives $\sum_{i\in I} f(z_i) - m_i = v(I)$ or, due to the budget balance constraint, $\sum_{i\in I} f(z_i) = v(I)$. Thus, if the payoff vector for an outcome (z, m) is in the core, the allocation of goods, z, must maximize the social welfare, subject to the conservation of goods constraint. In addition, the payments must be such that no coalition of agents has incentive to not accept the allocation of goods z and payment, and exchange only within their coalition.

The value $v(I)$ is the maximum social welfare that can be achieved by redistribution of the initial endowments. The optimization problem defining $v(I)$ is convex. A solution z^* to the optimization problem is an allocation of goods that maximizes the total welfare. In other words, z^* is said to be *efficient*. To gain insight, let us consider the problem in detail and identify the dual optimization problem. Let $p \in \mathbb{R}^\ell$ denote a vector of multipliers associated with the conservation of goods constraint, $\sum_{i\in I} z_i = \sum_{i\in I} w_i$. The entries of p can be interpreted as unit prices for the goods, in which case a payoff vector for the agents is given by $(f_i(z_i^*) - p(z_i^* - w_i))_{i\in I}$, where $p(z_i - w_i)$ is the inner product of p and $(z_i - w_i)$ (i.e. $p^T(z_i - w_i)$).

The optimization problem defining $v(I)$ is the following, which we call the primal problem:

$$\begin{aligned} \max_{z\geq 0} \ & \sum_{i\in I} f_i(z_i) \\ \text{subject to } & \sum_{i\in I} z_i = \sum_{i\in I} w_i. \end{aligned}$$

This convex optimization problem satisfies the Slater condition if $\sum_{i\in I} w_i > 0$ coordinatewise, in which case strong duality holds. The Lagrangian function is

$$L(z, p) = \sum_{i\in I} f_i(z_i) - \sum_{i\in I} p(z_i - w_i).$$

The dual optimization problem is

$$\min_{p\in\mathbb{R}^\ell} \left(\sum_{i\in I} \max_{z_i\geq 0} \{f_i(z_i) - p(z_i - w_i)\} \right). \tag{7.8}$$

A saddle point pair for this primal-dual pair of problems, (p^*, z^*), is called a *competitive equilibrium* (or Walrasian equilibrium). Equivalently, (p^*, z^*) is a competitive equilibrium if and only if:[2]

- z^* is a feasible allocation: $z^* \geq 0$ and $\sum_{i\in I} z_i^* = \sum_{i\in I} w_i$.
- $z_i^* \in \arg\max_{z_i\geq 0}\{f_i(z_i) - p^*(z_i - w_i)\}$ for each $i \in I$.

[2]This follows from Theorem 1.35 if the functions f_i are continuously differentiable.

The payoff vector x^* for a competitive equilibrium (p^*, z^*) is given by $x_i^* = f_i(z_i^*) - p^*(z_i^* - w_i)$. If the functions f_i are strictly concave then the competitive equilibrium is unique. Also in that case, by the envelope theorem, Proposition 6.21, the gradient of the i^{th} term of the sum in (7.8) with respect to p is $z_i(p) - w_i$, so the condition $\sum_{i\in I} z_i^* = \sum_{i\in I} w_i$ implies the gradient of the dual objective function at p^* is zero.

A competitive equilibrium is an equilibrium for *price-taking* agents, and the prices are adjusted so that the supply of each good matches the demand. To expand on this point, suppose agent i is presented with the vector of unit prices p^* for the various goods, and has the option of selecting a vector $z_i \in \mathbb{R}_+^\ell$ representing how much of each good the agent will end up with, by possibly selling some of the goods in his/her initial endowment w_i and possibly buying some other goods. The condition $z_i^* \in \arg\max_{z_i \geq 0}\{f_i(z_i) - p^*(z_i - w_i)\}$ means that z_i^* is the best response for agent i, under the assumption that the price vector p^* does not depend on the agent's choice of z_i. The feasibility condition for the competitive equilibrium means the price vector p^* is adjusted so that, under the assumption of price taking agents, the total supply of each good is equal to the demand.

The opposite of a price-taking agent is a strategic agent, who takes into account the impact of his/her choice of z_i on the price charged. Such strategic behavior of an agent is rational if the agent controls a significant portion of the market, in other words, if the agent has significant market power. The notion of competitive equilibrium, in contrast, is most relevant for large markets, with a large number of agents, such that the choice of each individual agent has only a small impact on any other agent.

Proposition 7.16 *(Payoff vectors for competitive equilibria are in the core) The payoff vector for any competitive equilibrium of the market is in the core of the induced cooperative game. In particular, the core is nonempty.*

Proof. Let $S \subset I$. It must be shown that $\sum_{i\in S} x_i^* \geq v(S)$, where v is defined by (7.7). Equivalently, by the definition of the x_i^*'s and v, it suffices to show that for any $(z_i)_{i\in S}$ with $\sum_{i\in S} z_i = \sum_{i\in S} w_i$,

$$\sum_{i\in S} f_i(z_i^*) - p^*(z_i^* - w_i) \geq \sum_{i\in S} f(z_i).$$

However, this follows from the fact that for each i, $f_i(z_i^*) - p^*(z_i^* - w_i) \geq f(z_i) - p^*(z_i - w_i)$, and the fact $\sum_{i\in S} p^*(z_i - w_i) = 0$ (which follows from $\sum_{i\in S} z_i = \sum_{i\in S} w_i$). ■

Example 7.17 *(Example of a market M with transferable payments) $I = \{1,2\}$, $\ell = 1$, $w_1 = w_2 = 1$, $f_1(z) = \ln z$ and $f_2(z) = 4\ln z$. We shall find the core and the competitive equilibrium. First, the coalitional valuation function v is identified. As always, $v(\emptyset) = 0$. If there is only one agent then no transfer of goods is possible, so $v\{1\} = f_1(w_1) = \ln(1) = 0$ and $v\{2\} = f_2(w_2) = 4\ln(1) = 0$. Finally,*

$$v\{1,2\} = \max_{z_1,z_2\geq 0, z_1+z_2\leq 2} (\ln z_1 + 4\ln z_2) = \ln(0.4) + 4\ln(1.6) \approx -0.9163 + 1.8800 = 0.9637.$$

The core constraints are $x_i \geq v\{i\} = 0$ for $i \in \{1,2\}$, and $x_1 + x_2 = v\{1,2\}$, so, geometrically speaking, the core is the line segment in $\mathbb{R}^2$ with endpoints $(v\{1,2\}, 0)$ and $(0, v\{1,2\})$. The efficient allocation is $(z_1^*, z_2^*) = (0.4, 1.6)$, which entails the transfer of 0.6 units of good from agent 1 to agent 2. The competitive equilibrium price p^* is given by $p^* = f_i'(z_i^*) = \frac{1}{0.4} = \frac{4}{1.6} = 2.5$. Under p^*, agent 2 pays $(0.6)(2.5) = 1.5$ to agent 1. The payoff vector is thus $x^* = (\ln(0.4) + 1.5, 4\ln(1.6) - 1.5) \approx (0.6836, 0.3800)$, which is a point in the core, as required by Proposition 7.16.

A clean model for large markets is to start with an exchange market $M = (I, \ell, (w_i), (f_i))$ and an integer k and consider the market kM obtained by k-fold replication of M. Then I indexes the types of the agents in kM. The market kM has k agents of each type, w_i is the endowment of goods of each type i agent, and $f_i : \mathbb{R}_+^\ell \to \mathbb{R}_+$ is the production/utility/happiness function for each type i agent. The market kM is not the same as k-fold replication of the corresponding cooperative game defined above, because even if a coalition S in kM has more than one agent of a given type, all the agents in the coalition can exchange goods among themselves.

Given a coalition S in kM, let y_i denote the number of agents of type i in S. Then $y = (y_i)_{i\in I} \in \{0, \ldots, k\}^I$. Since all agents of the same type have the same production function, $v(S)$ is determined by y, so we can define $\gamma(y)$ to equal $v(S)$. In particular, the grand coalition of M, I, corresponds to $y = \mathbf{1}$, were $\mathbf{1}$ is the vector of all ones, so $v(I) = \gamma(\mathbf{1})$. Let $\overline{x} \in \mathbb{R}^I$, and for $k \geq 1$ let $\overline{x}^{(k)}$ denote the payoff vector for kM such that each agent of type i gets payoff $\overline{x}_i$. By abuse of notation, we say that $\overline{x}$ is a payoff vector for kM and it is in the core of kM if $\overline{x}^{(k)}$ is in the core of kM. With this convention, the value of the grand coalition in kM is $kv(I) = k\gamma(\mathbf{1})$, and if (p^*, z^*) is a competitive equilibrium for M with payoff vector x^*, then (p^*, z^*) can also be considered to be a competitive equilibrium for kM with payoff vector x^*. In that case, z_i^* is considered to be the final good profile for an agent of type i, and x_i^* is the payoff of an agent of type i.

Vectors of the form $\overline{x}^{(k)}$ respect the principle of equal pay for equal contribution. For many M, the core of the kM contains payoff vectors of other types. However, if there is a vector x' in the core of kM, and if $\overline{x}$ corresponds to paying agents of each type i the average of $x'_{i'}$ over the k agents i' of type i, then $\overline{x}$ (or, equivalently, $\overline{x}^{(k)}$) is also in the core. To verify this, let S be an arbitrary coalition for kM with corresponding vector y, so that y_i is the number of agents of type i in S for each $i \in I$. Let S' be a coalition with the same descriptor y, such that for each type i, the coalition S' includes the y_i agents of type i with the smallest payments under x'. Then $\sum_{i'\in S} \overline{x}_{i'}^{(k)} \geq \sum_{i'\in S'} x'_{i'} \geq v(S') = v(S)$, so that $\overline{x}$ is in the core of kM as claimed.

Example 7.18 *(Let M be the market of Example 7.17. Let's find the vectors $\overline{x} = (\overline{x}_1, \overline{x}_2)$ in the core of $2M$ (the k-fold market kM for $k = 2$). The calculations in Example 7.17 imply $\gamma(0,0) = \gamma(1,0) = \gamma(0,1) = 0$ and $\gamma(1,1) = 0.9637$, giving the constraints $\overline{x}_i \geq 0$ and $\overline{x}_1 + \overline{x}_2 \geq 0.9637$. Since $\gamma(2,2) = 2\gamma(1,1)$, the feasibility constraint for $2M$ is $2\overline{x}_1 + 2\overline{x}_2 = \gamma(2,2)$, which is the same as the feasibility constraint for M: $\overline{x}_1 + \overline{x}_2 = 0.9637$. There are two types of coalitions in $2M$ not in M, corresponding to $y = (2,1)$ and $y = (1,2)$, consisting of three agents: two of one type and one of another type. We can assume without loss of optimality that agents of the same type get the same final allocation of goods (e.g. if z'_{11} and z'_{12} are the profiles of goods obtained by two agents of type 1, they can be assumed to be equal). Therefore,*

$$
\begin{aligned}
\gamma(2,1) &= \max_{z'_{11}, z'_{12}, z'_2 \geq 0 : z'_{11} + z'_{12} + z'_2 = 3} \ln z'_{11} + \ln z'_{12} + 4 \ln z'_2 \\
&= \max_{z_1, z_2 \geq 0 : 2z_1 + z_2 = 3} 2(\ln z_1) + 4 \ln z_2 \\
&= 2 \ln 1/2 + 4 \ln 2 = 2 \ln 2 \approx 1.3863
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
\gamma(1,2) &= \max_{z_1, z_2 \geq 0 : z_1 + 2z_2 = 3} \ln z_1 + 2(4 \ln z_2) \\
&= \ln 1/3 + 8 \ln 4/3 \approx 1.2028.
\end{aligned}
$$

Thus, in addition to the constraints $\overline{x}_i \geq 0$ and $\overline{x}_1 + \overline{x}_2 = 0.9637$, there are the constraints $2\overline{x}_1 + \overline{x}_2 \geq 1.3836$ and $\overline{x}_1 + 2\overline{x}_2 \geq 1.2028$. These constraints imply the core of $2M$ is given by $0.4199 \leq \overline{x}_1 \leq 0.7246$ and $\overline{x}_2 = 0.9637 - \overline{x}_1$. The competitive equilibrium $(0.6836, 0.3800)$ for M, is also the competitive equilibrium for kM for all $k \geq 1$, and it is in the core of $2M$ as required.

Example 7.18 suggests that as $k \to \infty$, the core of kM shrinks to the set of payoff vectors for competitive equilibria. Such result was conjectured in Edgeworth (1881) and proved in Debreau and Scarf (1963) and Aumann (1964). A version is given here.

Proposition 7.19 *(Core shrinking to competitive equilibrium payoffs as $k \to \infty$) Let $\overline{x} \in \mathbb{R}^I$. Then $\overline{x}$ is in the core of kM for all $k \geq 1$ if and only if $\overline{x}$ is the payoff vector of a competitive equilibrium for the original market M.*

Proof. To avoid details involving subgradients, we give a proof under the added assumption that the functions f_i are strictly concave, in addition to being continuous and increasing. By definition, for any $y \in \{0, \ldots, k\}^I$,

$$\gamma(y) = \max \sum_{i \in I} y_i f_i(z_i) \tag{7.9}$$

$$\text{over } z = (z_i)_{i \in I}, z_i \in \mathbb{R}^\ell, \text{ with } z \geq 0$$

$$\text{subject to } \sum_{i \in I} y_i z_i = \sum_{i \in I} y_i w_i,$$

which by strong convex duality yields:

$$\gamma(y) = \min_{p \in \mathbb{R}^\ell} \sum_{i \in I} y_i \max_{z_i \geq 0} \{f_i(z_i) - p(z_i - w_i)\}. \tag{7.10}$$

The representation (7.10) shows that γ is a concave function, and by the envelope theorem, γ is continuously differentiable and $\nabla\gamma(y) = x^*(y)$, where $x^*(y)$ is the payoff vector for the saddle point solution $(p^*(y), z^*(y))$ of (7.9) and (7.10).

By definition, a vector $\overline{x} \in \mathbb{R}^I$ is in the core of kM if and only if $\gamma(y) \leq y^T \overline{x}$ for all $y \in \{0, \ldots, k\}^I$, with equality for $y = k\mathbf{1}$, i.e., $\gamma(k\mathbf{1}) = k\mathbf{1}^T\overline{x}$. The function γ is homogeneous of degree one. In other words, for any $\alpha > 0$, $\gamma(\alpha y) = \alpha\gamma(y)$. Therefore, $\overline{x} \in \mathbb{R}^I$ is in the core of kM if and only if $\gamma(\mathbf{1}) = \mathbf{1}^T\overline{x}$ and $\gamma(y) \leq y^T\overline{x}$ for all $y \in \cup_{\alpha>0}\alpha\{0, \ldots, k\}^I$. The union over all $k \leq 1$ of the sets $\cup_{\alpha>0}\alpha\{0, \ldots, k\}^I$ is dense in $\mathbb{R}^I_+$, so if $\overline{x}$ is in the core of kM for all $k \leq 1$, $\gamma(y) \leq y^T\overline{x}$ for all $y \in \mathbb{R}^I_+$, with equality at $y = \mathbf{1}$. In other words, $\overline{x}$ is a supergradient, and hence the gradient, of γ at $y = \mathbf{1}$. Therefore, $\overline{x} = \nabla\gamma(\mathbf{1}) = x^*(\mathbf{1})$, which means $\overline{x}$ is the payoff vector of a competitive equilibrium of the base game corresponding to $k = 1$. ■

7.3 The Shapley value

Return to the general setting of a cooperative game (I, v). An alternative to considering the core is to somehow assign a feasible payoff vector $(x_i)_{i \in I}$ to any such game, where, as before, the vector x is defined to be feasible if $\sum_{i \in I} x_i = v(I)$. The Shapley value profile, often simply called the Shapley value, can be defined using the following notation.

Let $\mathcal{P}_n$ denote the set of $n!$ permutations of I

For $\pi \in \mathcal{P}_n$, let $S_i(\pi) = \{j \in I : j \text{ is before } i \text{ in } \pi\}$. (Note that $i \notin S_i(\pi)$.)

For a coalition S and player i with $i \notin S$, let $\triangle_i(S) = v(S \cup \{i\}) - v(S)$, which is the marginal value of adding i to S.

The Shapley value profile, or Shapley value, $(x_i)_{i\in I}$, is given by:

$$x_i \triangleq \frac{1}{n!}\sum_{\pi\in\mathcal{P}_n} \triangle_i(S_i(\pi)).$$

In words, x_i is the expected marginal value of adding player i to the set of players already added, given that the grand coalition is built up one player at a time, with the order of the players being chosen uniformly at random.

Remark 7.20 *(a) Proposition 7.8 states the following (using somewhat different notation): If v is supermodular ($v(S\cup T)+v(S\cap T)\geq v(S)+v(T)$) then for any permutation π, the payoff profile $(x_i)=(\triangle_i(S_i(\pi)))_{i\in I}$, is in the core. Since the core is a convex set, the Shapley value is thus also in the core if v is supermodular.*

(b) Shapley showed the mapping $v \xrightarrow{F} x^{Shapley}$ for fixed I is the unique mapping subject to three axioms for a given set I:

- *linear: $F(v_1+v_2)=F(v_1)+F(v_2)$.*
- *symmetric with respect to permutation of the agents*
- *no value for dummies: $F_i(v)=0$ if $v(S)=v(S\cup\{i\})$ for every coalition S not including agent i.*

See [18] for a proof.

Example 7.21 *Consider a majority coalition formed in a parliament, consisting of the union of three disjoint blocks, s, m and ℓ, having 20 seats, 30 seats, and 50 seats, respectively, Suppose the threshold for a majority is 65 seats. Thus, the value of the majority coalition is one, and if only the large block and either the small or medium block were in the coalition, it would still have value one. To compute the Shapley value profile for blocks within the majority coalition, consider all 3! ways the majority coalition could have been built up by having the blocks join one at a time.*

π	$(\triangle_s(S_s(\pi)),\triangle_m(S_m(\pi)),\triangle_\ell(S_\ell(\pi)))$
$sm\ell$	$(0,0,1)$
$s\ell m$	$(0,0,1)$
$ms\ell$	$(0,0,1)$
$m\ell s$	$(0,0,1)$
ℓsm	$(1,0,0)$
ℓms	$(0,1,0)$

So $x_{Shapley}=\left(\frac{1}{6},\frac{1}{6},\frac{4}{6}\right)$.

For sake of comparison, let's find the core of the game. The constraints for (x_s,x_m,x_ℓ) to be in the core are $x_s,x_m,x_\ell\geq 0$, $x_s+x_m+x_\ell=1$, $x_s+x_\ell\geq 1$ and $x_m+s_\ell\geq 1$. The only solution is $(0,0,1)$; the core is the singleton set, $\{(0,0,1)\}$. Although the core is not empty, the Shapley value is not in the core. (The game is not supermodular, or else this would be a contradiction.)

Bibliography

[1] B.-T. Aharon and A. Nemirovski. Lecture notes optimization III. https://www2.isye.gatech.edu/~nemirovs/OPTIII_LectureNotes2018.pdf, Spring 2013.

[2] R. Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[3] M. Benaim. A dynamical system approach to stochastic approximations. *SIAM Journal on Control and Optimization*, 34(2):437–472, 1996.

[4] I. Bomze. Non-cooperative two-person games in biology: A classification. *International journal of game theory*, 15(1):31–57, 1986.

[5] V. Borkar. *Stochastic approximation: a dynamical systems viewpoint*. Springer, 2009.

[6] J. Cassels. *Economics for mathematicians*, volume 62. Cambridge University Press, 1981.

[7] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[8] D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[9] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

[10] D. Fudenberg and J. Tirole. *Game Theory (MIT Press)*. MIT Press, 11th printing edition, 1991.

[11] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

[12] D. Kreps and R. Wilson. Sequential equilibria. *Econometrica: Journal of the Econometric Society*, pages 863–894, 1982.

[13] I. Menache and A. Ozdaglar. Network games: Theory, models, and dynamics. *Synthesis Lectures on Communication Networks*, 4(1):1–159, 2011.

[14] P. Milgrom and I. Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

[15] D. Monderer and L. Shapley. Potential games. *Games and Economic Behavior*, pages 124–143, 1996.

[16] R. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[17] R. Myerson. *Game Theory: Analysis of Conflict*. First Harvard University Press, 1991.

[18] M. Osborne and A. Rubinstein. *A course in game theory*. MIT press, 1994.

[19] J. Rosen. Existence and uniqueness of equlibrium points for concave n-person games. *Econometrics*, 33, 1965.

[20] H. L. Royden. *Real Analysis*. Prentice Hall, 3 edition, 1988.

[21] J. Shamma and G. Arslan. Unified convergence proofs of continuous-time fictitious play. *IEEE Trans. on Automatic Control*, pages 1137–1141, 2004.

[22] Y. Shoham and K. Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[23] R. Wilson. Game theoretic analysis of trading processes. In T. Bewley, editor, *Advances in Economic Theory*. Cambridge University Press, 1987.

[24] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proc. Intnl Conf Machine Learning (ICML-03)*, pages 928–936, 2003.

Index