Dynamics of Learning and Iterated Games

Lecture Notes

MATH60007/70007/97069

Sebastian van Strien (Imperial College)

January 26, 2024

Contents

0 Introduction

This module aims to describe is how people, animals, plants or computers can learn over time. We will cover models for:

- The evolution (i.e. learning) of populations. In other words, to understand how genetic mutations which improve the performance of a population evolve over time. Such models are also often used in economics.
- An explanation why in spite of the common assumption that all actions are aimed at maximising an individuals' payoff (and so are based on selfish motivations), altruistic behaviour naturally evolves in nature.
- How learning works which is based on reinforcing behaviour which repeats actions which led to good payoff. Such models are widely used in the computer science literature, and are also used by for example Deep Mind (a subsidiary of Google) when they developing technology which allows computers to learn to play games (or to solve other problems). Here we will focus on a setting where several players have different interests, rather than on a purely probabilistic setting.
- How learning works which is based on no-regret learning. There the idea is that a human, or a computer evaluates whether they would have done better in the past if alternative actions had been taken. If so, then this is used to inform future decisions.

As you will see from these notes, we will try to understand the foundations of the dynamical aspects of learning. However, two of the three projects by which this module will be examined are rather practical and may include a significant amount of computer coding.

0.1 What is this course about?

The notion of Nash equilibrium aims to describe how players optimise their behaviour in a competitive environment. For this reason it is prevalent in many areas of science: economics, biology, engineering etc.

The aim of this course is to highlight some situations where the notion of Nash equilibrium, or related notions, are given a more dynamic interpretation. So a Nash equilibrium would be a stationary point of some differential equation, or of some other dynamical process.

Example 0.1. Consider a population of birds where some will always fight about a grain (let us call these hawks or hawkish birds), and others will always do some posturing but then retreat rather than fight (dove-like). The payoff of getting the grain is G, and the price for getting hurt is $-C$. We assume that $0 < G < C$.

If a hawk bird meets another hawk bird, one wins (and gets payoff G) and the other looses (and get payoff $-C$). On average this means the payoff for each is $(G + -C)/2$. If a dove meets a dove, then one will get the grain but the other will not get hurt. In this way we get the following payoff matrix

$$\begin{array}{rc} \text{meeting} & \begin{array}{cc} \text{Hawk} & \text{Dove} \end{array} \\ \begin{array}{r} \text{payoff to Hawk} \\ \text{payoff to Dove} \end{array} & \begin{pmatrix} \frac{G-C}{2} & G \\ 0 & \frac{G}{2} \end{pmatrix} \end{array}.$$

How is it that not the entire species develops hawkish behaviour?
Suppose that the frequency of 'hawkish' birds in the population is x and 'dove-like' birds is $1-x$. Then the average 'fitness' is

$$\begin{array}{ll} x(G-C)/2+(1-x)G & \text{for hawkish birds} \\ x\cdot 0+(1-x)G/2 & \text{for dove-like.} \end{array}$$

If $x=1$ then the fitness of hawkish birds is <0 and of dove-like birds is $=0$, and so the number of dove-like birds will increase and the number of hawk-like will decrease. (Hawk-like birds are constantly fighting and getting injured, whereas the dove-like will occasionally get lucky.) When $x=0$, the fitness is G resp. $G/2$, so the number of hawk-like birds will increase. Equality holds when $x=G/C$.

Example 0.2 (Prisoner dilemma)**.** Consider two prisoners, each in a separate rooms so that they cannot communicate. The prisoners get a higher reward by betraying the other (defecting), but if both cooperate (so stay silent) they get a reduced sentence. For example we may have the following situation:

$$\begin{array}{rc} \text{Prisoner II} & \text{Coop} \quad \text{Defects} \\ \begin{array}{r}\text{Pris. I Coop} \\ \text{Pris I Defects}\end{array} & \begin{pmatrix} -1,-1 & -3,0 \\ 0,-3 & -2,-2 \end{pmatrix}. \end{array}$$

This table describes the payoff (the number of years prison sentence) in various scenarios. For example if prisoner II defects but prisoner I cooperates, then prisoner II will be released and prisoner I will be 3 years in prison. What should the prisoners do? If II cooperates then I is better off to defect (he then gets 0 years rather than 1 year prison sentence). If II defects then he still better to defect (he gets 2 years rather than rather than 3 years). The same holds for II. So the rational behaviour is for both prisoners to defect, resulting in a prison sentence of two years for each.

Example 0.3 (Repeated prisoner dilemma / repeated donation game)**.** What if the previous set-up is repeated every year? Or what if two players are asked every week to make a donation of £5 and if they do the other player gets a donation of £15, otherwise nothing. So the situation is described by

$$\begin{array}{rc} \text{II} & \text{donates} \quad \text{declines} \\ \begin{array}{r}\text{I donates} \\ \text{I declines}\end{array} & \begin{pmatrix} 10,10 & -5,15 \\ 15,-5 & 0,0 \end{pmatrix}. \end{array}$$

Of course if this game took place only one week, then this is again a prisoner dilemma game. If this play is repeated many times then the considerations of the players will change of course. We will discuss this situation in this course. (A political scientist called Axelrod, even organises computer tournaments which explore which strategy is the most optimal. One strategy is called TFT (Tit for Tat).)

To emphasise that it is important to consider the detailed set-up of the game, let us consider the following:

Example 0.4 (Parrondo paradox)**.** Consider two games Game A and Game B:

- In Game A, you lose £1 every time you play.
- In Game B, you count how much money you have left. If it is an even number, you win £3. Otherwise you lose £5.

Say you begin with £100. If you start playing Game A exclusively, you will obviously lose all your money in 100 rounds. Similarly, if you decide to play Game B exclusively, you will also lose all your money in 100 rounds.

However, consider playing the games alternatively, starting with Game B, followed by A, then by B, and so on (BABABA...). It should be easy to see that you will steadily earn a total of £2 for every two games.

Thus, even though each game is a losing proposition if played alone, because the results of Game B are affected by Game A, the sequence in which the games are played can affect how often Game B earns you money, and subsequently the result is different from the case where either game is played by itself.

Different types of game dynamics

In this course we will consider various types of game dynamics.

- **Replicator dynamics**, both for one population and then for two players (or two populations). This kind of dynamics is often considered in biology and in economics and is related to Darwin's idea of 'survival of the fittest'. One question we will try to answer is how it is possible that one has *altruistic behaviour*, even in these models.
- **Iterated prisoner dilemma games**, where two players have to repeatedly make a decision whether to corporate or to defect, but where they do not know how long for. In such a setting, the range of strategies the players can choose is much larger, and this can lead to quite unexpected behaviour.
- **No-regret learning** A different variant of a learning algorithm is that of no-regret learning. This is based on the idea that if a different action in the past would have given a better payoff, assuming the other player would have done the same, then different decisions should be made in the future.
- **Reinforcement learning** The notion of payoff to players also leads to various learning principles: the higher the payoff from a given action is, the more likely this action will be taken in the future. There are various models which make this intuitive notion precise. This and the next learning algorithm are both used heavily by IT companies such as DeepMind, but are also studied by engineers, economists and so on.
- **Best response dynamics and fictitious play**, which was introduced in economics and game theory as a dynamics which was meant to converge to the Nash equilibria. This turned out to be not the case, but this dynamics is still often studied.

0.2 Prerequisites

No other background will be required in this module than what is covered in any differential equations course; no background is required in game theory.

0.3 Practical Arrangements

This module will be taught using the flipped class room model. This means that

- Students will be expected to read weekly approximately 10 pages of these lecture notes before the class takes place in which the corresponding material is covered using exercises.
- Each section in the notes contains one or more exercises which will test whether you have understood the material. The 10-12 Wednesday class will be dedicated to these exercises and to Q&A sessions.
- The 3-4 Tuesday class will be dedicated to a more general overview of the material.
- The course will be examined by project. The arrangements and support for this is outlined in the next subsection.

0.4 Assessment arrangment

- This course will be examined by a project, together with a presentation on this project. The possible topics for this project will be handed out after a few weeks into the term. This project will need to be submitted at the end of week 1 of term 2.
- Slots will be offered around week 6 and 7 to discuss your choice of topics, and further optional slots to discuss your progress with the project.

0.5 References

These lecture notes are fully self-contained. If you want to read further, the main references for these lectures are the following books, which are both available online through Imperial's library:

- Hofbauer & Sigmund, *Evolutionary games and population dynamics*
- Sigmund, *The Calculus of Selfishness*.

For completeness I will also list references by chapter:

- Chapter 1 is about **replicator dynamics** in one-player games (so a population where genes compete against each other). The books by Hofbauer & Sigmund, *Evolutionary games and population dynamics* and by Weibull, *Evolutionary game theory* are standard references.
- Chapter 2 is about **two player replicator games**. More on the classification of 2×2 replicator dynamics can be found in Hofbauer & Sigmund, *Evolutionary games and population dynamics* but a more detailed description can be found in chapter 3 of Cressman, *Evolutionary Dynamics and Extensive Form Games*. The description of a chaotic replicator dynamics system is given in Sato, Akiyama and Crutchfield, *Stability and diversity in collective adaptation*, Physica D, 210, 2015, 21-57.
- Chapter 3 is about **repeated prisoner dilemma games**. More can be found in Sigmund, *The Calculus of Selfishness*.
- Chapter 4 is about **regret learning**, and follows Sergiu Hart and Andreu Mas-Colell, *A simple adaptive procedure leading to correlated equilibrium*, Econometrica, Vol. 68, No. 5, 2000., 1127-1150. This algorithm is widely used in the AI community.
- Chapter 5 is about **reinforcement learning**, but in the game theoretic setting. Section 5.1 follows essentially Posch, *Cycling in a stochastic learning algorithm for normal form games*, J Evol Econ (1997) 7: 193-207. But there is an extensive literature on this.

  Some of this work is grounded in the field of behavioural economics - which aims model how people learn, e.g. Erev & Roth, *Predicting how people play games: Reinforcement learning in experimental games with unique*, mixed strategy equilibria. Amer. Econ. 1998, Rev. 88, 848-881.

  Section 6.2 follows Posch *Cycling in a stochastic learning algorithm for normal form games*, Evolutionary. Economics (1997) 7, 193-207 and Beggs, *On the convergence of reinforcement learning*, Journal of Economic Theory (2005) 122, 1-36.

  For a discussion on approximating discrete 'random' dynamical systems by differential equations can be found in for example Benaïm, *Dynamics of stochastic approximation algorithms*, in: Seminaire de Probabilité,XXXIII, Lecture Notes in Mathematics, vol. 1709, Springer, Berlin, 1999, https://doi.org/10.1073/pnas.1109672110 it is shown that for a very large class of games and for large class of learning dynamics one has complicated dynamics.
- Chapter 6 is about **best response dynamics**. More on this can be found in Hofbauer, *Deterministic Evolutionary Game Dynamics*, Proceedings of Symposia in Applied Mathematics Volume 69, 2011. Shapley was the first to observe that there is a periodic orbit in the RPS game (which sometimes is called a Shapley game):

  - Shapley, *Some Topics in Two-Person Games*, in M. Dresher, L. S. Shapley and A. W. Tucker, eds., Advances in Game Theory, Annals of Mathematics Studies No. 52, 1-28, 1964.

  For results on the bifurcations of periodic orbits and chaotic best response dynamics of the generalised Shapley game, see `http://wwwf.imperial.ac.uk/~svanstri/publications_by_subject.php` and specifically

  - Colin Sparrow, SvS & Christopher Harris, Fictitious Play in 3x3 Games: the transition between periodic and chaotic bahavior. Games and Economic Behavior 63, (2008), 259-291.
  - Colin Sparrow & SvS, Fictitious Play in 3x3 Games: chaos and dithering behaviour, Games and Economic Behavior 73 (2011), 262-286.

- Chapter 7 discusses another version of best response dynamics, namely **fictitious games**. Amongst other things, this chapter explores whether the limit sets of best response dynamics (and of the replicator dynamics) have game theoretic properties. Sections 7.1 up to Section 574 follow Ostrovski & van Strien, *Payoff performance of fictitious play, Journal of Dynamics and Games*, vol 1, issue 4, October 2014. In Ostrovski & van Strien, *Payoff performance of fictitious play, Journal of Dynamics and Games*, vol 1, issue 4, October 2014 it is shown that the average payoff for **both** players is often better if they play (FP) than if they play (NE). It would be interesting to explore whether this is also true for the other learning dynamics considered in these lecture notes (or specifically for the systems considered by Galla & Farmer).

  Section 7.5 follows J. Hofbauer, S. Sorin and Y. Viossat (2009) *Time average replicator and best reply dynamics*. Math. Operations Res. 10 (2), 263–269.

- More general references for the chapters on learning are: Fudenberg & Levine, *The Theory of Learning in Games*. MIT Press. (1999) and Young, *Strategic learning and Its limits*, Oxford, U.K, (2004), or from the machine learning point of view, see for example Nisan, Roughgarden, Tardos and Vazirani, *Algorithmic Game Theory*, 2007.

1 Replicator dynamics for one population

1.1 Nash equilibrium of one population

We consider a large population where each individual can have one of a finite set of n pure strategies. You might think of these as individuals which can have one of n different traits (e.g. colour of eyes, fighting behaviour, personality characteristics, opinions etc.) Let x_i denote the frequency of strategy i in the population. So $(x_1, \dots, x_n)$ is a probability vector. Let $\Delta_n = \{x \in \mathbb{R}^n; 0 \le x_i \le 1, x_1 + \cdots + x_n = 1\}$ be the $(n-1)$-dimensional simplex. So $(x_1, \dots, x_n) \in \Delta_n$. Usually we will fix n and write Δ. For later use, let e_i be the vector in Δ with a coefficient 1 on the i-th coordinate.

Let us assume consider a populations in which an invader who chooses strategy i against a strategy j receives payoff a_{ij}. Assuming the populations uses a mixed strategy $(y_1, \dots, y_n)$, with random matching (that is, random encounters) this leads to the following **linear payoff** to an invader choosing strategy (or action) i:

$$a_i(y) = \sum_j a_{ij} y_j = (Ay)_i$$

where A is the matrix (a_{ij}). If the invader uses a mixed strategy x this gives a payoff

$$\text{Payoff}(x, y) := x \cdot Ay.$$

A probability vector $\hat{x} \in \Delta$ is called a **Nash equilibrium (NE)** iff

$$x \cdot A\hat{x} \le \hat{x} \cdot A\hat{x}, \forall x \in \Delta. \tag{1.1}$$

and a **strict Nash equilibrium** if

$$x \cdot A\hat{x} < \hat{x} \cdot A\hat{x}, \forall x \in \Delta \text{ with } x \neq \hat{x}. \tag{1.2}$$

Note that $x \cdot A\hat{x} \le \hat{x} \cdot A\hat{x}, \forall x \in \Delta$ means that the invader cannot do better by choosing anything other than $\hat{x}$. We say that $\hat{x}$ is a **pure** NE if $\hat{x} = e_i$ for some i.

An equivalent way of formulating the notion of Nash equilibrium is to define the **best response** map

$$\mathcal{B}R(x) = \arg\max_{y \in \Delta} y \cdot Ax \stackrel{def}{=} \{y' \in \Delta; y \cdot Ax \le y' \cdot Ax \ \forall y \in \Delta\}.$$

Then $\hat{x}$ is a NE iff $\hat{x} \in \mathcal{B}R(\hat{x})$.

Example 1.1. Consider a game determined by $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. What are its Nash equilibria? To see this, note that

$$\mathcal{B}R(x) = \begin{cases} e_1 & \text{if } x_1 > x_2 \\ e_2 & \text{if } x_1 < x_2 \\ \Delta & \text{if } x_1 = x_2 \end{cases}$$

So $e_i \in \mathcal{B}R(e_i)$ and $\mathcal{B}R\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \ni \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ and $x \notin \mathcal{B}R(x)$ for any other vector. So e_1, e_2 and $z := (e_1 + e_2)/2 := \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ are the Nash equilibria. Note that $Az = z$ and so $x \cdot Az = 1/2$ for **each** $x \in \Delta$. So z is not a strict NE. On the other hand, e_1, e_2 are both strict NE.

Example 1.2. Consider a game determined by $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. What are its Nash equilibria? Note that Δ in this case is a triangle. $\mathcal{B}R$ takes values e_1, e_2, e_3 on three convex regions, see Figure 1, which meet at $(1/3, 1/3, 1/3)$. At this midpoint, one has that $(1/3, 1/3, 1/3) \in \mathcal{B}R(1/3, 1/3, 1/3) = \Delta$ and so this is a NE. Taking $Z_{1,2}$ to be the line-segment connecting $(1/3, 1/3, 1/3)$ to the midpoint between e_1 and e_2 we have $\mathcal{B}R(x) =< e_1, e_2 >$ for all $x \in Z_{1,2}$ and so where $Z_{1,2}$ intersects $\partial\Delta$ we get another NE. Continuing this analysis, we see there are precisely 7 NE's.

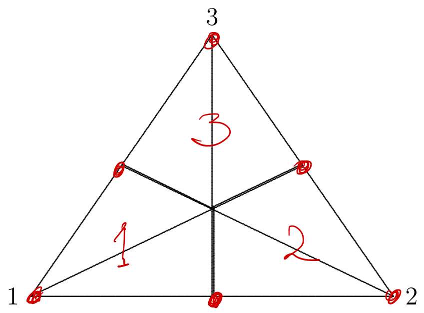


Figure 1: The indifference lines, the best response regions and the Nash equilibria corresponding to Example 1.2.

Exercise 1.1. 1. Give a real life example in which you clarify the notion of Nash equilibrium.

2. Consider the game determined by $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. What are its Nash equilibria? Compute the lines $Z_{ij} = \{x \in \Delta; (Ax)_i = (Ax)_j\}$. What is the relationship between the NE and these lines?

1.2 Evolutionary stable strategies

$\hat{x}$ is an **evolutionary stable equilibrium (ESS)** if for all $x \in \Delta, x \neq \hat{x}$ one has for $\epsilon > 0$ small enough,

$$x \cdot A(\epsilon x + (1-\epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x}). \tag{1.3}$$

Here the size of $\epsilon > 0$ is allowed to depend on x.

Lemma 1.1. strict NE $\implies$ ESS $\implies$ NE.

Remark: As we will prove at the end of this chapter, every game has a Nash equilibrium. On the other hand, there are games without an ESS.

Proof. First assume $\hat{x}$ is a strict NE. Then $x \cdot A\hat{x} < \hat{x}A\hat{x}$. This inequality is what the ESS condition (1.3) reduces to if we take $\epsilon = 0$. By continuity the ESS condition then also holds for $\epsilon > 0$ small.

Now assume that x is an ESS. For each $x \neq \hat{x}$ we can let $\epsilon \to 0$ in the ESS condition and we obtain $x \cdot A\hat{x} \leq \hat{x}A\hat{x}$. $\square$

Example 1.3. Consider a game determined by $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. What are its ESS? We know this matrix has three NE's. Let us check which of these are ESS's. First consider whether $z = (e_1 + e_2)/2$ is an ESS. Note that $x \cdot Az = 1/2$ for each $x \in \Delta$. For z to be an ESS, we need that for $\epsilon > 0$ small, and all $x \neq z$ we have $x \cdot A(\epsilon x + (1-\epsilon)z) < z \cdot A(\epsilon x + (1-\epsilon)z)$. This reduces to $x \cdot Ax < z \cdot Ax$. This is supposed to hold for all $x \neq z$ so in particular for $x = e_1$, but this is clearly not true. So z is not an ESS. Note that, in fact, we could have used the next lemma to conclude that z is not an ESS.

Let us now show that e_1 is an ESS. So we need to show that for all $x = (x_1, x_2) \neq \hat{e}_1$ and all $\epsilon > 0$ sufficiently small, $x \cdot A(\epsilon x + (1-\epsilon)e_1) < e_1 \cdot A(\epsilon x + (1-\epsilon)e_1)$ when $x \neq e_1$. This is equivalent to $\epsilon(x_1^2 + x_2^2) + (1-\epsilon)x_1 < \epsilon x_1 + (1-\epsilon)$ which holds for $\epsilon > 0$ small since $x_1 \neq 1$. [If we take $\epsilon = 0$, then the ESS inequality becomes $x_1 = x \cdot Ae_1 < e_1 \cdot Ae_1 = 1$ which clearly holds when $x \neq e_1$. So for $\epsilon > 0$ small the ESS inequality also holds.] In the same way we get that e_2 is also an ESS.

Lemma 1.2. If $\hat{x}$ is a Nash equilibrium then there exists $c \in \mathbb{R}$ so that $(A\hat{x})_i = c$ for each i for which $\hat{x}_i > 0$. In particular, $c = \hat{x} \cdot A\hat{x}$. Moreover, $\hat{x} \in \Delta$ is so that there exists $c \in \mathbb{R}$ with $(A\hat{x})_i = c$ for each i then $\hat{x}$ is a NE.

Finally, if $\hat{x} \in \operatorname{int} \Delta$ is an ESS, then there exists no other NE.

Proof. Suppose $\hat{x}$ is a NE. Then substituting e_i for x in the definition (1.1) of NE, implies

$$e_i \cdot A\hat{x} \leq \hat{x} \cdot A\hat{x}.$$

This holds for all $i = 1, \dots, n$. Write $\hat{x} = \sum \lambda_i e_i$ with $\lambda_i \geq 0$ and $\sum \lambda_i = 1$. Summing over the previous inequality we get

$$\hat{x} \cdot A\hat{x} = \sum \lambda_i e_i \cdot A\hat{x} \leq \sum \lambda_i \hat{x} \cdot A\hat{x} = \hat{x} \cdot A\hat{x}.$$

But we would obtain here a strict inequality if $e_i \cdot A\hat{x} < \hat{x} \cdot A\hat{x}$ for some i for which $\lambda_i > 0$, which is clearly impossible. Hence

$$e_i \cdot A\hat{x} = \hat{x} \cdot A\hat{x} \text{ for all } i = 1, \dots, n \text{ for which } \hat{x}_i > 0.$$

proving the first assertion of the lemma. If $(A\hat{x})_i = c$ for all i then $\mathcal{B}R_A(\hat{x}) = \Delta$ and so $\hat{x}$ is a NE. The 2nd assertion in the lemma follows.

From the first assertion, it follows that if $\hat{x}$ is an interior NE, then for each $x \in \Delta$ one has $x \cdot A\hat{x} = c$ (here we use that x is a probability vector). Assume that $\hat{x}$ is also an ESS, i.e. that for each $x \neq \hat{x}$ and $\epsilon > 0$ small one has $x \cdot A(\epsilon x + (1-\epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x})$. But since $x \cdot A\hat{x} = \hat{x} \cdot A\hat{x} = c$, this inequality reduces to $x \cdot Ax < \hat{x} \cdot Ax$ for each $x \neq \hat{x}$. So any $x \neq \hat{x}$ cannot be a NE. $\square$

Lemma 1.3. The ESS assumption (1.3) is equivalent to the assumption that for all $y \neq \hat{x}$ sufficiently close to $\hat{x}$,

$$y \cdot Ay < \hat{x} \cdot Ay \tag{1.4}$$

Moreover, if $\hat{x} \in \operatorname{int} \Delta$ is an ESS then

$$y \cdot Ay < \hat{x} \cdot Ay \text{ for all } y \in \Delta \tag{1.5}$$

Proof. First associate to $\hat{x}$ a compact set $\Lambda \subset \partial\Delta$ so that $\hat{x} \notin \Lambda$ and so that for each $y \neq \hat{x}$ the line $l(t) = (1-t)\hat{x} + ty$, $t \geq 0$ intersects Λ. [1]

By the ESS inequality (1.3) for each $x \in \Lambda$ there exists $\epsilon_0(x) > 0$ so that $x \cdot A(\epsilon x + (1-\epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x})$ for all $\epsilon \in (0, \epsilon_0(x))$. There exists an open neighbourhood $U(x)$ so that the same inequality also holds for all $x' \in U(x)$ (replacing x' by x). By compactness of Λ, the open cover $\cup_{x\in\Lambda} U(x)$ has a finite sub-cover. It follows that there exists $\epsilon_0 > 0$ so that

$$x \cdot A(\epsilon x + (1-\epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x})$$

for each $x \in \Lambda$ and each $\epsilon \in (0, \epsilon_0)$.

By the choice of Λ for each $y \neq \hat{x}$ sufficiently close to $\hat{x}$ can be written in the form $y = (1-\epsilon)\hat{x} + \epsilon x$ for some $x \in \Lambda$ and for some $\epsilon \in (0, \epsilon_0)$. Substituting the definition of y in the previous displayed equation gives

$$x \cdot Ay < \hat{x} \cdot Ay.$$

Multiplying this inequality by ϵ and adding to both sides the inequality the term $(1-\epsilon)\hat{x} \cdot Ay$, gives the required inequality $y \cdot Ay < \hat{x}Ay$.

Now assume that $\hat{x} \in \operatorname{int}\Delta$ is an ESS. Then by the previous lemma, there exists c so that for all x, $x \cdot A\hat{x} = c = \hat{x} \cdot A\hat{x}$. This means that $x \cdot A(\epsilon x + (1-\epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x})$ reduces to the required inequality $x \cdot Ax < \hat{x} \cdot Ax$ for all $x \in \Delta$. $\square$

Example 1.4. Consider the payoff matrix $A = \begin{pmatrix} \frac{G-C}{2} & G \\ 0 & \frac{G}{2} \end{pmatrix}$ considered in the introduction. For simplicity take $G = 2$ and $C = 4$ so that $A = \begin{pmatrix} -1 & 2 \\ 0 & 1 \end{pmatrix}$. So $Ax = \begin{pmatrix} -x_1 + 2x_2 \\ x_2 \end{pmatrix}$ and this gives $\mathcal{B}R(x) = e_1$ if $x_2 > x_1$ and $\mathcal{B}R(x) = e_2$ if $x_2 < x_1$. Furthermore, for $\hat{x} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ we have $\mathcal{B}R(x) = \Delta$ and so $\hat{x}$ is a NE. This is not a strict NE because $A\hat{x} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \hat{x}$ and therefore $xA\hat{x} = \hat{x}A\hat{x}$ for all x. To check that it is an ESS consider $y = \hat{x} + \begin{pmatrix} \epsilon \\ -\epsilon \end{pmatrix}$. Since $A\hat{x} = \hat{x}$ we have $Ay = \hat{x} + \begin{pmatrix} -3\epsilon \\ -\epsilon \end{pmatrix}$. Hence

$$y \cdot Ay = \left(\hat{x} + \begin{pmatrix} \epsilon \\ -\epsilon \end{pmatrix}\right) \cdot \left(\hat{x} + \begin{pmatrix} -3\epsilon \\ -\epsilon \end{pmatrix}\right) = (1/2) - 2\epsilon - 2\epsilon^2$$

while

$$\hat{x} \cdot Ay = \hat{x} \cdot \left(\hat{x} + \begin{pmatrix} -3\epsilon \\ -\epsilon \end{pmatrix}\right) = (1/2) - 2\epsilon$$

and therefore by Lemma 1.3 $\hat{x}$ is an ESS.

[1] Let us show how one can choose this set $\Lambda \subset \partial\Delta$. If $\hat{x}$ is in the interior of Delta then take $\Lambda = \partial\Delta$. If $\hat{x}$ is in the interior of one the sides of the triangle, then take Λ to be equal to the union of the other sides. So if for example $\hat{x}$ is in the interior of (e_1, e_2) and not on of the corner points, then you take $\Lambda = [e_1, e_3] \cup [e_3, e_2]$. If $\hat{x}$ is a corner point then take the the opposite side. So if for example $\hat{x} = e_1$ then take $\Lambda = [e_2, e_3]$. This choice for Λ ensures that each point y in the simplex is a convex combination of $\hat{x}$ and a point in Λ and so the role of Λ is to pick 'directions'. The reason for choosing such a set Λ is because we need to use some compactness argument which gives the existence of the required $\epsilon > 0$. Each point y near $\hat{x}$ corresponds to a point $x \in \Lambda$ (just take the half-line from $\hat{x}$ through y, and vice versa each point $x \in \Lambda$ corresponds to a half-line and so direction along which a point y near $\hat{x}$ can lie. Corresponding to each half-line through $\hat{x}$ there exists a suitable $\epsilon > 0$ so that provided y is on that half-line and is $\epsilon_0(x)$ close to $\hat{x}$ the required inequality holds.

Example 1.5. Let us show that $A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$ does not have any ESS. To start with let us determine its NE's. Assume $x \in \operatorname{int} \Delta$ is a NE. Then there exists c so that $(Ax)_i = c$ for each i. So $x_2 - x_3 = -x_1 + x_3 = x_1 - x_2 = c$, which gives $c = 0$ and $x_i = 1/3$, and the point $(1/3, 1/3, 1/3)$ is a NE. Now consider the set

$$Z_{ij} = \{x; (Ax)_i = (Ax)_j\}$$

of x so that the i and j-th coordinate of Ax are the same. These can be computed relatively easily (see lecture).

From this diagram it follows that for each i the set of $x \in \Delta$ for which $e_i \in \mathcal{B}R(x)$ is an non-empty convex region containing $(1/3, 1/3, 1/3)$ and one of the corners of the triangle (see the lectures for a drawing). This diagram also implies that $(1/3, 1/3, 1/3)$ is the only NE.

Is $\hat{x} = (1/3, 1/3, 1/3)$ a ESS? Again we need to consider the inequality $x \cdot A(\epsilon x + (1 - \epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x})$. This reduces to $x \cdot Ax < \hat{x} \cdot Ax$. That is, $x_1(x_2 - x_3) + x_2(-x_1 + x_3) + x_3(x_1 - x_2) < (1/3)[(x_2 - x_3) + (-x_1 + x_3) + (x_1 - x_2)]$. Note that both the left and right hand side are zero, so the inequality does NOT hold and so $\hat{x}$ is not an ESS. It follows that this games has no ESS.

Exercise 1.2.

1. Determine the NE for $A = \begin{pmatrix} 0 & 2 & -1 \\ -1 & 0 & 2 \\ 2 & -1 & 0 \end{pmatrix}$. Does this game have a strict NE or an ESS?

2. Show that e_1, e_2, e_3 are ESS points for $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

1.3 Replicator dynamics

One proposal to describe a mechanism which explains why Nash equilibria and ESS can appear as a dynamic process is the following system of differential equations

$$\dot{x}_i = x_i((Ax)_i - x \cdot Ax), i = 1, \ldots, n. \tag{1.6}$$

This means that one considers $x_i(t)$ is a real-valued variable depending on time t, rather than as a rational number.

Note that if $x(t)$ is a probability vector then $\sum_i \dot{x}_i = \sum_i x_i((Ax)_i - x \cdot Ax) = x \cdot Ax - x \cdot Ax = 0$. Moreover, if $x_i(0) = 0$ then $x_i(t) = 0$ for all t. It follows that if $x(0)$ is a probability vector then $x(t)$ is a probability vector for all t. Hence by a fundamental theorem from the theory of differential equations, solutions of (1.6) exist for all time $t \in \mathbb{R}$.

The rationale behind the differential equation (1.6) is the following: let the population be divided up in n types, and let x_i be the proportion of type i (so that x is a probability vector). Then $\dot{x}_k/x_i$ describes the growth rate of type i. The replicator equation assumes that $\dot{x}_i/x_i$ is equal to the fitness $(Ax)_i = e_i \cdot Ax$ of this type i *minus* the mean average fitness $\sum_i x_i \cdot (Ax)_i = x \cdot Ax$ of the population. In particular if $x_i > 0$ and $(Ax)_i > x \cdot Ax$ then $\dot{x}_i > 0$.

Note that (1.6) implies that

$$\frac{d}{dt}\frac{x_i}{x_j} = \frac{x_i}{x_j}((Ax)_i - (Ax)_j). \tag{1.7}$$

Lemma 1.4 (Nash equilibria and equilibria of the ODE)**.** .

1. Any Nash equilibrium $\hat{x}$ is an equilibrium of the replicator equation (1.6).
2. If $\hat{x} \in \text{int}\,\Delta$ is an equilibrium of (1.6) then $\hat{x}$ is a NE.
3. If $\hat{x}$ is Lyapounov stable, then it is a NE. ("Lyapounov stable" is defined in the Appendix.)

Proof. By Lemma 1.2, if $\hat{x}$ is a Nash equilibrium then there exists a c so that $(A\hat{x})_i = c$ for each i for which $\hat{x}_i > 0$. It follows that $(A\hat{x})_i - \hat{x} \cdot A\hat{x} = 0$ for each of such i. For the other i's one has $\hat{x}_i = 0$. It follows that $\hat{x}$ is a zero of (1.6), proving part (1) of the lemma.

If $\hat{x} \in \text{int}\,\Delta$ is an equilibrium of (1.6) then $\hat{x}_i > 0$ for all i and so $(A\hat{x})_i = \hat{x} \cdot A\hat{x}$ for each i. So there exists c so that $(A\hat{x})_i = c$ for all i. By Lemma 1.2 this implies that $\hat{x}$ is a NE.

If $\hat{x}$ is not a Nash equilibrium then there exists x so that $x \cdot A\hat{x} > \hat{x} \cdot A\hat{x}$. It follows that there exists i so that $e_i \cdot A\hat{x} > \hat{x} \cdot A\hat{x}$. Hence there exists $\epsilon > 0$ so that for x close to $\hat{x}$ (here we reuse the name x), $(Ax)_i - x \cdot Ax = e_i \cdot Ax - x \cdot Ax > \epsilon$. Hence $\dot{x}_i > \epsilon x_i$ when x is close to $\hat{x}$ and so it is impossible that $x(t) \to \hat{x}$ as $t \to \infty$. □

Example 1.6. Give an example of a system for which not every stationary point $\hat{x}$ is a NE. (Hint: there may be indices i with $\hat{x}_i = 0$ when $(A\hat{x})_i > c$ where c is as above.)

Example 1.7. Describe what happens for

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.8}$$

What are its NE's? What are its ESS's?
Let us first check whether A has any interior NE x. Then $(Ax)_i = c$ and so $x_2 = 2x_3 = x_3 = c$. So $c = x_3 = x_2 = 0$. So there is no interior NE.
To see whether there are other NE's we consider the set $Z_{i,j} = \{x \in \Delta; (Ax)_i = (Ax)_j\}$ where the best response is indifferent between i, j. These sets $Z_{1,2} = \{x_2 = 2x_3\}$, $Z_{1,3} = \{x_2 = x_3\}$ and $Z_{2,3} = \{2x_3 = x_3\} = \{x_3 = 0\}$ are all lines through e_1. By considering the position, of these lines, it follows the triangle contains regions with non-empty interior where $\mathcal{B}R = e_1$ and $\mathcal{B}R = e_2$, see Figure 2 below. Here we use that $Ax = \begin{pmatrix} x_2 \\ 2x_3 \\ x_3 \end{pmatrix}$. It follows that e_1 is the unique NE.
Now we can ask whether $\hat{x} = e_1$ is a EES? Note that $A\hat{x} = 0$, so $x \cdot A(\epsilon x + (1-\epsilon)\hat{x}) < \hat{x} \cdot A(\epsilon x + (1-\epsilon)\hat{x})$ reduces to $x \cdot Ax < \hat{x} \cdot Ax$ which is equivalent to $x_1 x_2 + x_2 2x_3 + x_3 x_3 < x_2$. which is not the case when $x_1 = x_2 = 0, x_3 = 1$ (or when $x_1 = 1, x_2 = x_3 = 0$), and so e_1 is not an ESS.
By considering $(x_2/x_3)' = (x_2/x_3)(2x_3 - x_3)$, $(x_1/x_2)' = (x_1/x_2)(x_2 - 2x_3)$, $(x_1/x_3)' = (x_1/x_3)(x_2 - x_3)$ we see that on the sides of the triangle there are no additional singularities of the flow, except in the corners. Moreover, it follows that the phase portrait is as in Figure 2 . Note that each of the corners e_i is a singularity, but only e_1 is a Nash equilibrium.

Exercise 1.3. 1. Consider the replicator dynamics associated to the following system $A = \begin{pmatrix} 0 & 10 & 1 \\ 10 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. What are the singularities and the ESS points.

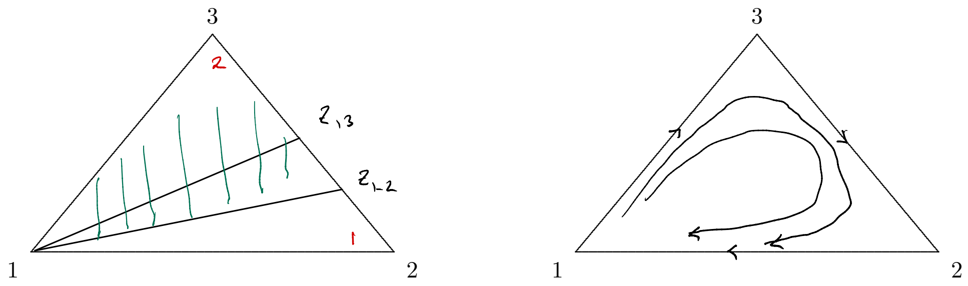


Figure 2: The indifference lines and the flow corresponding to Example 1.7.

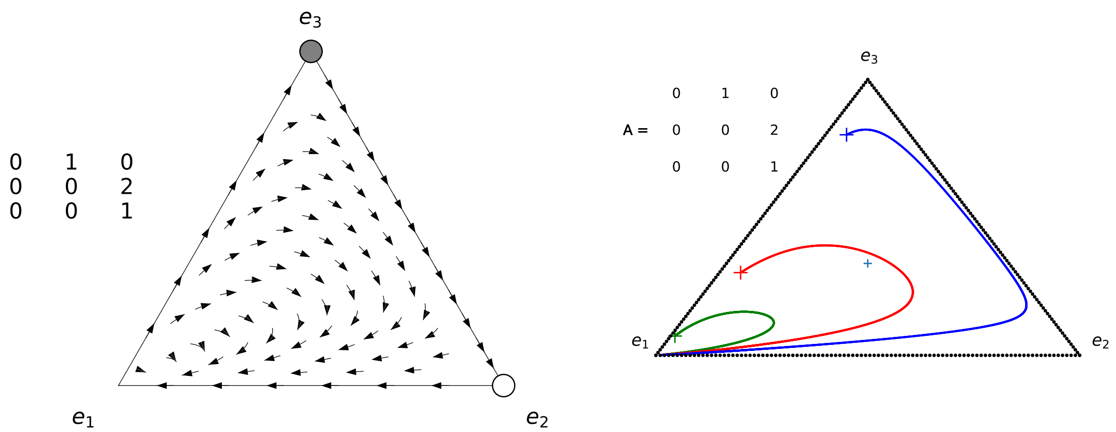


Figure 3: The arrow plot and a computer drawn plot of the flow corresponding to Example 1.7.

2. What is the effect to the replicator dynamics $\dot{x}_i = x_i((Ax)_i - x \cdot Ax)$ of adding to the first column of A the vector $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. What are the NE and the ESS for this system?

3. Why do solutions of the replicator equations starting at some $x(0) \in \Delta$ exist for all $t \in \mathbb{R}$. (Hint: use a theorem from your course on differential equations.)

1.4 ESS points are asymptotically stable for the replicator system

We say that $X \subset \mathbb{R}^n$ is *convex* if $t \cdot x + (1-t)y \in X$ for all $x, y \in \mathbb{R}^n$ and for all $t \in [0,1]$. Let us say that a function $f \colon X \to \mathbb{R}$ is *convex* if $f(t \cdot x + (1-t)y) \le tf(x) + (1-t)f(y)$ for all $x, y \in \mathbb{R}^n$ and for all $t \in [0,1]$ and it is called *concave* if the opposite inequality holds throughout.

Theorem 1.1. If $\hat{x}$ is an ESS then it is asymptotically stable for the replicator system.
If $\hat{x} \in \operatorname{int}\Delta$ is an ESS then it globally attracts all initial points $x \in \operatorname{int}\Delta$.

Proof. Consider the function $P(x) = \prod x_i^{\hat{x}_i}$. Let us show that this has a unique maximum at $\hat{x}$. First notice that when f is a convex function on some interval I, then $f(\sum p_i x_i) \le \sum p_i f(x_i)$ for $x_1, \dots, x_n \in I$ and all p_i with $p_i \ge 0$ and $\sum p_i = 1$. If f is strictly convex, then a strict inequality holds except when all the x_i are equal. Applying this to $f = \log$ on $[0, \infty]$ (which is concave, so we get the opposite inequality) gives $\sum \hat{x}_i \log(\frac{x_i}{\hat{x}_i}) = \sum_{\hat{x}_i > 0} \hat{x}_i \log(\frac{x_i}{\hat{x}_i}) \le \log \sum_{\hat{x}_i > 0} x_i \le \log \sum x_i = \log 1 = 0$. Hence $\sum_i \hat{x}_i \log x_i \le \sum_i \hat{x}_i \log \hat{x}_i$ and so $P(x) \le P(\hat{x})$ with equality only if $x = \hat{x}$.

So let us now show that we can consider P as a Lyapounov function:

$$\frac{\dot{P}}{P} = \frac{d}{dt}(\log P) = \frac{d}{dt} \sum \hat{x}_i \log x_i = \sum_{\hat{x}_i > 0} \hat{x}_i \frac{\dot{x}_i}{x_i} =$$

$$= \sum \hat{x}_i((Ax)_i - x \cdot Ax) = \hat{x} \cdot Ax - x \cdot Ax$$

Since by assumption $\hat{x}$ is an ESS, the equation (1.4) gives that the r.h.s. is > 0 and so $\dot{P} > 0$ for all $x \neq \hat{x}$ close to $\hat{x}$. It follows that orbits starting near $\hat{x}$ converge to $\hat{x}$.

If $\hat{x} \in \operatorname{int}\Delta$ then (1.5) implies that $\dot{P}/P > 0$ everywhere and so $\hat{x}$ attracts all points in $\operatorname{int}\Delta$. $\square$

Example 1.8. Consider the matrix $A = \begin{pmatrix} 0 & 6 & -4 \\ -3 & 0 & 5 \\ -1 & 3 & 0 \end{pmatrix}$. Show that $E = (1/3, 1/3, 1/3)$ is a rest point which is asymptotically stable. To see this, compute the eigenvalues of the linearisation at this fixed point. Show that this point is not an ESS, by showing that $e_1 = (1, 0, 0)$ is an ESS.

Solution: $Ax = \begin{pmatrix} 6x_2 - 4x_3 \\ -3x_1 + 5x_3 \\ -1x_1 + 3x_2 \end{pmatrix}$ and so the lines $Z_{i,j}$ all go through E. From this one can see that the lines $Z_{i,j}$ are as in the figure, and so E is a Nash equilibrium. This also determines the singularities and the arrows on the sides of the triangle, as $(x_i/x_j)' = (x_i/x_j)[(Ax)_i - (Ax)_j]$. Indeed, $Z_{2,3} \cap [e_2, e_3]$ and $Z_{1,3} \cap [e_1, e_3]$ are singularities, and of course e_1, e_2, e_3 are also singularities. Note that 2 and 3 are suboptimal strategies at $Z_{2,3} \cap [e_2, e_3]$ and so this point is not a Nash equilibrium. Similarly, e_2, e_3 are not Nash equilibria. On the other hand, $Z_{1,3} \cap [e_1, e_3]$ and e_1 are Nash equilibria. In summary, this game has three Nash equilibria and three additional singularities.

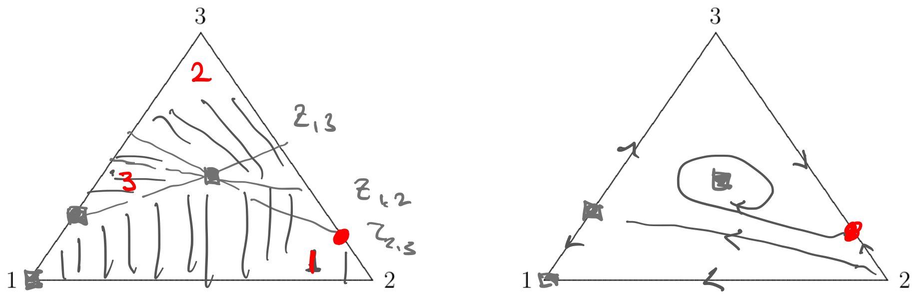


Figure 4: The indifference lines and the flow corresponding to Example 1.8.

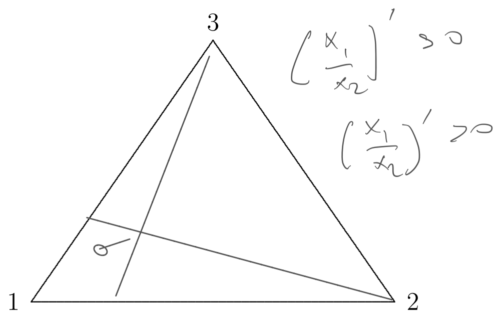


Figure 5: By computing $(x_1/x_2)'$ and $(x_1/x_3)'$ at some point you can determine which direction the flow points towards, see the text in Example 1.8.

The singularity $Z_{2,3} \cap [e_2, e_3] = (0, 5/8, 3/8)$ is a saddle point. Indeed on $[e_2 e_3]$ we have $(x_2/x_3)' = (x_2/x_3)[5x_3 - 3x_2]$. This shows that the arrows along this side point towards $(0, 5/8, 3/8)$. **Exercise: show that this point is indeed a saddle point.**

To compute the eigenvalues in $E = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$ we write $x_i = E + h_i$ and let h be the vector with components h_i. Note that since x and E are probability vectors, $\sum h_i = 0$ and so we can replace h_3 by $-h_1 - h_2$. Since all components of AE are equal we have that $h \cdot AE = 0$ and so

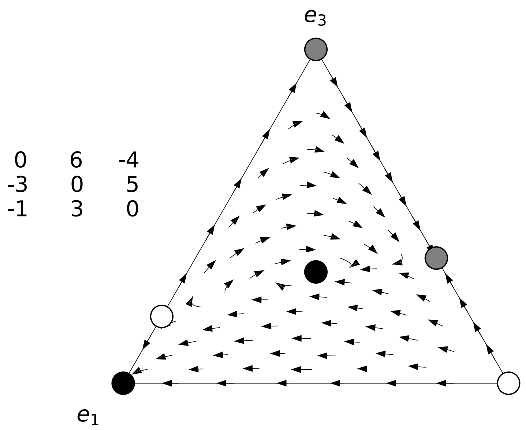
$$\begin{aligned} x \cdot Ax &= (E+h) \cdot A(E+h) \\ &= E \cdot AE + h \cdot AE + E \cdot Ah + O(h^2) \\ &= E \cdot AE + E \cdot Ah + O(h^2). \end{aligned} \tag{1.9}$$

and

$$(Ax)_i - x \cdot Ax = (Ah)_i - E \cdot Ah + O(h^2). \tag{1.10}$$

Taking $\mathbb{1}$ to be the vector with 1's we get

$$\begin{aligned} E \cdot Ah &= (1/3)\mathbb{1} \cdot Ah \\ &= (1/3)(-4h_1 + 9h_2 + h_3) \\ &= (-5/3)h_1 + (8/3)h_2 \end{aligned}$$



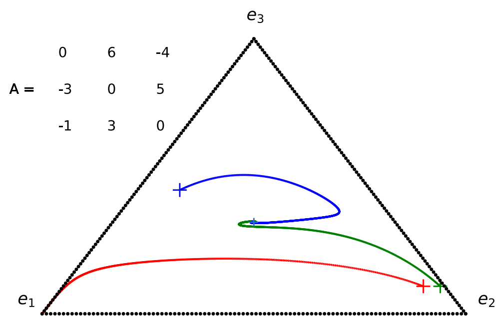


Figure 6: The arrow plot and a computer drawn plot of the flow corresponding to Example 1.8. To make sense of the flow diagram one needs to add carefully more initial conditions. This figure also does not show clearly that orbits spiral towards E. In that sense the hand-drawn flow drawn in Figure 4 shows more clearly what is going on.

and

$$\begin{aligned}(Ah)_1 &= 6h_2 - 4h_3 = 4h_1 + 10h_2,\\ (Ah)_2 &= -3h_1 + 5h_3 = -8h_1 - 5h_2.\end{aligned}$$

So $\dot{x}_i = x_i((Ax)_i - x \cdot Ax)$ gives

$$\begin{aligned}\dot{h}_1 &= ((1/3) + h_1)\,((17/3)h_1 + (22/3)h_2 + O(h^2))) =\\ &= (1/9)\,(17h_1 + 22h_2) + O(h^2).\\ \dot{h}_2 &= (1/9)\,(-19h_1 - 23h_2) + O(h^2).\end{aligned}$$

This implies that the linear part at E is equal to

$$(1/9)\begin{pmatrix} 17 & 22 \\ -19 & -23 \end{pmatrix}$$

The eigenvalues of the matrix are $(1/3)(-1 \pm i\sqrt{2})$.

To see that e_1 is an ESS it is sufficient to check that $(x - e_1) \cdot A(\epsilon x + (1-\epsilon)e_1) < 0$ when $\epsilon > 0$ small. Another way of seeing this, is to observe that it is sufficient to show that $P = x_1$ is a strict Lyapounov function. (To see that this is sufficient, have a look at the proof of the previous theorem. There it is shown that $\dot{P}/P = \hat{x} \cdot Ax - x \cdot Ax$ and by Lemma 1.3 ESS is equivalent to the statement that this term is positive for x close to $\hat{x}$.) But we have that $(x_1/x_3)' = (x_1/x_3)[(Ax)_1 - (Ax)_3]$ and $(x_1/x_2)' = (x_1/x_2)[(Ax)_1 - (Ax)_2]$ where the square bracket terms are both positive. This means that the speed vector along the line $P = x_1 = \epsilon$ lies in the cone in the figure, and so P is strictly increasing.

Additional arguments are needed to show that the saddle-separatrices are as shown in Figure 4.

Exercise 1.4. 1. Consider the function P from Theorem 1.1 taking $\hat{x}$ equal to e_1, e_2, e_3 and $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Draw the phase diagram for the replicator dynamics.

1.5 Further examples

Example 1.9. Consider the following matrix, determine the corresponding NE's and the phase diagrams of the replicator dynamics. $A = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}$;

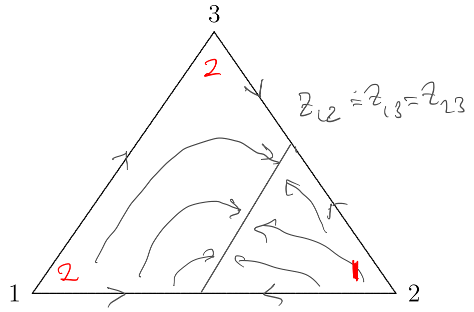


Figure 7: The indifference lines and the flow corresponding to Example 1.9.

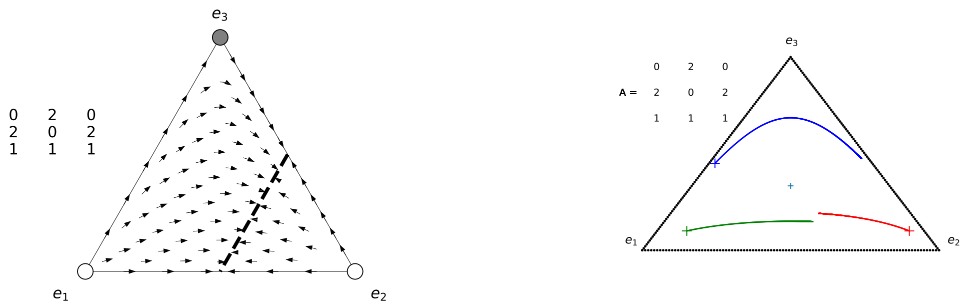


Figure 8: The arrow plot and a computer drawn plot of the flow corresponding to Example 1.9.

Solution: The corner points e_i are not NE's, as is clear from Figure 7. Next compute $Z_{1,2}$ and $Z_{1,3}$, which correspond to $2x_2 = 2x_1 + 2x_3$ resp $2x_1 + 2x_3 = x_1 + x_2 + x_3$. These are the same lines. $Z_{1,3}$ corresponds to $2x_2 = x_1 + x_2 + x_3 = 1$, so $x_2 = 1/2$. So $Z_{1,3} = Z_{1,2} = Z_{2,3}$ and this line consists entirely of NE's. Since each of these is a stationary point of the system, by Theorem 1.1, none of these points is an ESS. In summary, this system has infinitely many NE's, three additional singularities and no ESS.

The arrows along the boundary can be seen by using $(x_i/x_j)' = (x_i/x_j)[(Ax)_i - (Ax)_j]$. Along $[e_1, e_2]$ we get $(Ax)_1 - (Ax)_2 = (2x_2 - 2x_1)$, so a sign change at $x_1 = x_2 = 1/2$. Along $[e_1, e_3]$ we get $(Ax)_1 - (Ax)_3 = (2x_2 - 1) = -1 < 0$ and along $[e_2, e_3]$ we get $(Ax)_2 - (Ax)_3 = (2x_1 + 2x_3 - 1) = 2x_3 - 1$ which has a sign change. Along $Z_{1,2} = Z_{1,3}$ we have that $Ax = (1, 1, 1)$ so this means that all these points are singularities of the replicator system. Note that everywhere $(x_1/x_2)' = (x_1/x_2)((Ax)_1 - (Ax)_2) = (x_1/x_2)(2x_2 - (2x_1 + 2x_3)) = 4(x_1/x_2)(x_2 - 1/2)$ which shows that orbits converge to the line $Z_{1,2} = Z_{1,3} = Z_{2,3} = \{x_2 = 1/2\}$.

Example 1.10. Consider $A = \begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 5 \\ 5 & 0 & 4 \end{pmatrix}$ determine the corresponding NE's and the phase diagrams of the replicator dynamics. Is there an ESS?

To answer these questions we start by computing Z_{ij}: $Z_{1,2}$ corresponds to $x_1 + 4x_2 = 5x_3$, $Z_{1,3}$ to $5x_2 = 4x_1 + 4x_3$ and $Z_{2,3}$ to $x_2 + x_3 = 5x_1$. These lines intersect at $E := \hat{x} =$

$(3/18, 8/18, 7/18)$, so this is a Nash equilibrium. Note that from the form of the indifference equations, it follows that each side of Δ is intersected by precisely two indifference lines. This, and since $\mathcal{B}R(e_i) = e_{i-1}$, implies that there just two possible positions for the Z_{ij} lines, as shown in the figure. Since $Z_{1,2}$ does not intersect $[e_1 e_2]$, the situation is as in the left figure.

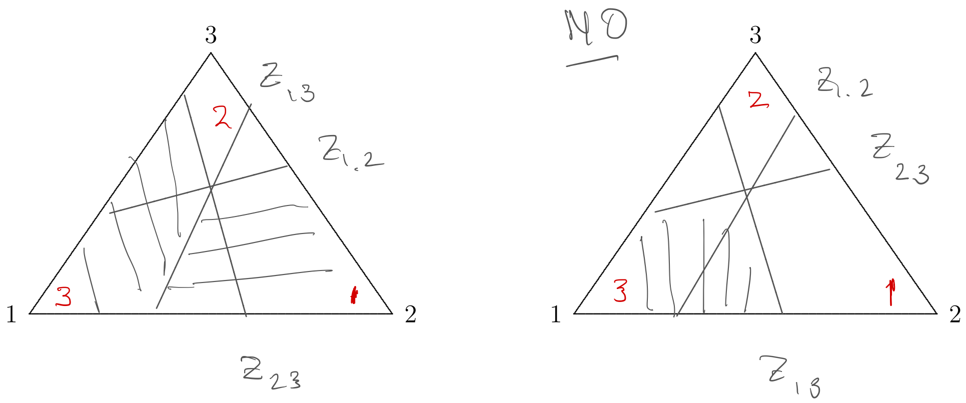


Figure 9: Two potential configurations of the indifference lines in Example 1.10. As explained in the text, the right configuration is impossible.

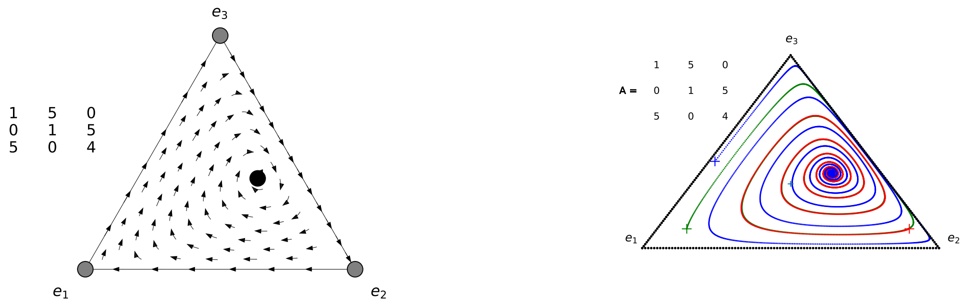


Figure 10: The arrow plot and a computer drawn plot of the flow corresponding to Example 1.10.

Once we see this, we also obtain that orbits are rotating about the NE. Is this NE an ESS? By Lemma 1.3 and since the NE lies in the interior of Δ, the ESS condition corresponds to $xAx < \hat{x}Ax$ for x close to the NE. Write $x = \hat{x} + h = (3/18 + h_1, 8/18 + h_2, 7/18 + h_3)$ where $h_1 + h_2 + h_3 = 0$. So we need to check $(x - \hat{x})Ax = (h_1 h_2 h_3)Ax < 0$. Since $\hat{x}$ is a NE and $h_1 + h_2 + h_3 = 0$, this is equivalent to $(h_1 h_2 h_3)A(h_1 h_2 h_3) < 0$. Since $h_1 + h_2 + h_3 = 0$, the last expression is equal to $h_1^2 + 5h_1h_2 + h_2^2 + 5h_2h_3 + 5h_1h_3 + 4h_3^2$. Substituting $h_3 = -h_1 - h_2$ gives that this is equal to

$$h_1^2 + 5h_1h_2 + h_2^2 - 5h_2h_1 - 5h_2^2 - 5h_1^2 - 5h_1h_2 + 4(h_1 + h_2)^2 = 3h_1h_2.$$

This expression does not have a constant sign for $h_1, h_2 \approx 0$. So the attracting NE is not an ESS. Nevertheless solutions converge to E. Indeed, write $x = \hat{x} + h$. Then, using the calculation from (1.10),

$$\begin{aligned} \dot{h}_1 &= (3/18 + h_1)[(h_1 + 5h_2) - \hat{x} \cdot Ah + O(h^2)] \\ \dot{h}_2 &= (8/18 + h_2)[(h_2 + 5h_3) - \hat{x} \cdot Ah + O(h^2)]. \end{aligned}$$

Note that $\hat{x}A = (19/9, 23/18, 34/9)$ and since $h_3 = -h_1 - h_2$ this gives $\hat{x}Ah = -(5/3)h_1 - (5/2)h_2$ and so we obtain

$$\begin{aligned} \dot{h}_1 &= (1/18)[(3h_1 + 15h_2) + (5h_1 + (15/2)h_2) + O(h^2)] \\ \dot{h}_2 &= (1/18)[(8h_2 - 40h_1 - 40h_2) + ((40/3)h_1 + 20h_2) + O(h^2)]. \end{aligned}$$

Or in simplified form:

$$\begin{aligned}\dot{h}_1 &= (1/18)[8h_1 + (45/2)h_2] + O(h^2)\\ \dot{h}_2 &= (1/18)[-(80/3)h_1 - 12h_2] + O(h^2)\end{aligned}$$

The linear part of this system is

$$\frac{1}{18}\begin{pmatrix} 8 & 45/2 \\ -80/3 & -12 \end{pmatrix}.$$

This has eigenvalues $-0.1111 \pm 1.2423i$ so the system is locally stable. (I've determined the eigenvalues using Matlab.) To show that the system is globally stable one needs additional methods.

Exercise 1.5. 1. Consider the replicator dynamics associated to the following system $A = \begin{pmatrix} 0 & 10 & 1 \\ 10 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ (see also Exercise 1.3(1)) and determine its phase portrait.

1.6 Rock-paper-scissor replicator game

There is a class of systems which have only one Nash equilibrium and for which $\mathcal{B}R(e_i) = e_{i+1}$ (or $\mathcal{B}R(e_i) = e_{i-1}$). So this suggests cyclic behaviour, and are therefore called rock-paper-scissor games. Let us consider the replicator dynamics in one example of this situation; in the general case the analysis is the same but computationally more involved.

Example 1.11. Consider the matrix $A = \begin{pmatrix} 0 & 1 & -b \\ -b & 0 & 1 \\ 1 & -b & 0 \end{pmatrix}$ when $b > 0$. If $b = 1$ then this is a zero-sum game because then $A + A^{tr} = 0$ (and we took $B = A^{tr}$ to be the payoff matrix of the 2nd player), but otherwise it is not a zero-sum game. Show that $V := x_1x_2x_3$ is a constant of motion (i.e. $t \mapsto V(x(t))$ is constant) when $b = 1$ and draw the phase diagram. If $b \neq 1$, it is a Lyapounov function; draw the phase diagram. This game is called a **rock-paper-scissor game**. These games will discussed again in Subsection 1.6.

Solution. Note that $\mathcal{B}R(e_i) = e_{i-1}$: so the best response is cyclic. This matrix has an interior NE at $(1/3, 1/3, 1/3)$. As in Theorem 1.1 take $P = (x_1x_2x_3)^{1/3}$. By the calculation in that theorem, $\dot{P}/P = \hat{x}\cdot Ax - x\cdot Ax = (\hat{x} - x)\cdot Ax$. Write $x = 1/3 + h_i$. A calculation shows that $(\hat{x}-x)\cdot Ax = (b/3-1/3)(h_1+h_2+h_3)+(b-1)(h_1h_2+h_1h_3+h_2h_3) = (1-b)(h_1^2+h_2^2+h_1h_2)$ where in the last equality we used that $h_3 = -h_1 - h_2$. So $\dot{P}/P > 0$ when $b \in (0,1)$ and $\dot{P}/P < 0$ when $b > 1$. So interior orbits starting at $x \neq E$, spiral out to the boundary when $b > 1$ and towards E when $b \in (0,1)$.

Let us see whether there are other Nash equilibria. This can be done in a number of ways. One way is to show that the indifference lines are as shown along the above figure. Along the boundary $x_2 = 0$, we have $Ax = (x_2 - bx_3, -bx_1 + x_3, x_1 - bx_2) = (-bx_3, -b + (1+b)x_3, 1 - x_3)$ where we used that along this boundary $x_1 = 1 - x_3$. Hence $\mathcal{B}R(e_1) = e_3$, $\mathcal{B}R(e_3) = e_2$ and $x \in Z_{2,3} \cap [e_1, e_3]$ along this side implies $x_3 = (1+b)/(2+b)$ and therefore $e_2 \cdot Ax = e_3 \cdot Ax > 0$. When $x \in Z_{1,3} \cap \{x_2 = 0\}$ then $x_3 = 1/(b-1) \notin [0,1]$ as $b > 0$ and similarly $x \in Z_{1,2} \cap \{x_2 = 0\}$ then $x_3 = b/(1+2b)$ and then $(Ax)_1 = (Ax)_2 < 0 < (Ax)_3$. So using the symmetry we obtain that the positions of Z_{ij} are as in Example 1.10.

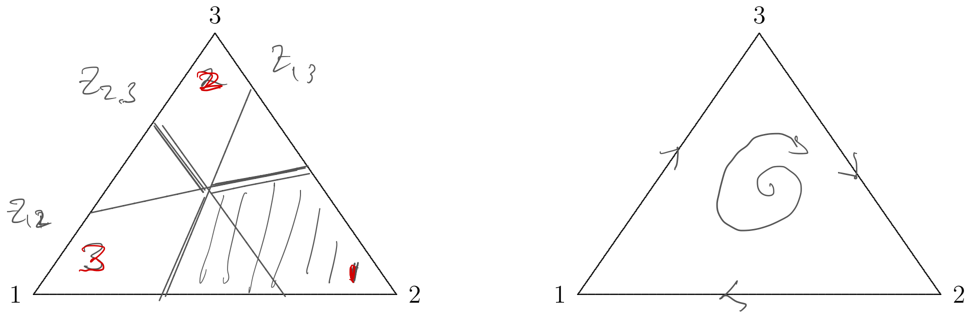


Figure 11: The rock-scissor-paper game from Example 1.11. On the right the situation where $b > 1$ is drawn.

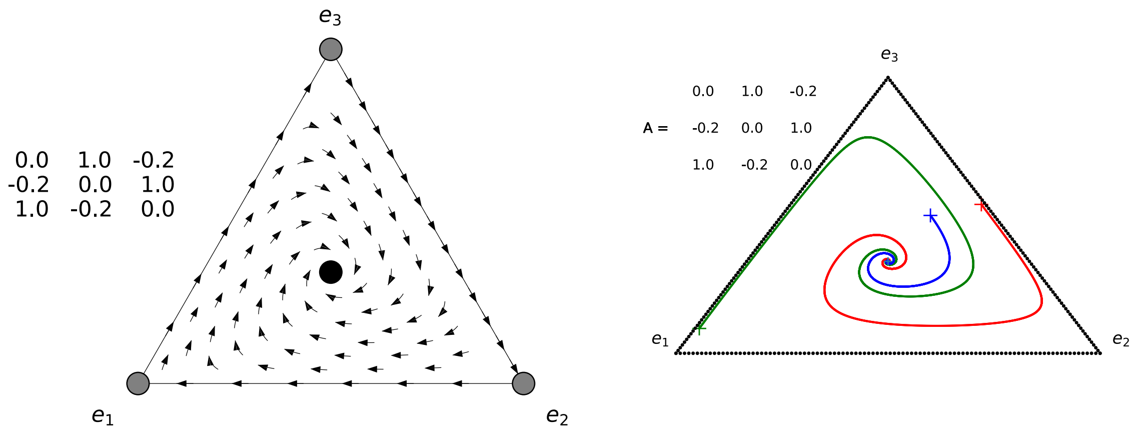


Figure 12: The arrow plot and a computer drawn plot of the flow corresponding to Example 1.11 for $b \in (0, 1)$.

Moreover, along $[e_1, e_3]$ one has $(Ax)_1 - (Ax)_3 = -bx_3 - (1 - x_3) = (1 - b)x_3 \leq 0$ as $b > 0$ and $x_3 \in [0, 1]$. So $(x_1/x_3)' < 0$ and there are no singularities along this boundary of the triangle. So solutions spiral out/in depending on whether $b > 1$ or $b \in (0, 1)$, and the arrows on the sides of Δ are as shown.

Lemma 1.5. Consider $A = \begin{pmatrix} 0 & 1 & -b \\ -b & 0 & 1 \\ 1 & -b & 0 \end{pmatrix}$ with $b > 1$. Then

$$z(T) = \frac{1}{T} \int_0^T x(t)\, dt \tag{1.11}$$

depends continuously on T and converges to some polygon with corners $A_1 = (1, b^2, b)/(1 + b + b^2)$, $A_2 = (b, 1, b^2)/(1 + b + b^2)$ and $A_3 = (b^2, b, 1)/(1 + b + b^2)$. You will be asked in Exercise 1.6 to show that A_i, A_{i+1}, e_{i+1} are collinear. (Later on we shall see that the triangle is the orbit under the so-called best response dynamics.)

Proof. Integrating the expression of the replicator dynamics and dividing by T gives

$$\frac{\log(x_i(T)) - \log(x_i(0))}{T} = \frac{1}{T} \sum_j a_{ij} \int_0^T x_j(t)\, dt - \frac{1}{T} \int_0^T x \cdot Ax\, dt. \tag{1.12}$$

Since $x(t)$ spends most of the time close to corners of the symplex (there the speed is small, and between corners it is large) and since $a_{ii} = 0$,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T x \cdot Ax\, dt \to 0. \tag{1.13}$$

(This statement is properly proved in the 2nd part of Exercise 1.6.)

Take a sequence $T_k \to \infty$ so that $x(T_k)$ converges to some point $w \in \partial\Delta$ with T_k chosen, to be definite, so to that $w \in (e_3, e_2)$, i.e. $w_1 = 0$ and so that $w_2, w_3 > 0$. Taking, if necessary, a subsequence of T_k we can assume that $z(T_k) = \frac{1}{T_k} \int_0^{T_k} x(t) dt$ converges to some point $z \in \Delta$.

Since $x_i(T_k) \to w_i \in (0, 1)$ for $i = 2, 3$ and since $T_k \to \infty$ we have

$$\frac{\log(x_i(T_k)) - \log(x_i(0))}{T_k} \to 0$$

for $i = 2, 3$ as $k \to \infty$. Hence, using (1.11), (1.12) and (1.13) we get that for $i = 2, 3$,

$$\frac{1}{T_k} \sum_j a_{ij} \int_0^{T_k} x_j(t)\, dt = (Az(T_k))_i \to 0$$

as $k \to \infty$. Since $z(T_k) \to z$ this implies $(Az)_2 = 0, (Az)_3 = 0$, i.e.

$$\sum_j a_{2j} z_j = \sum_j a_{3j} z_j = 0.$$

Using the definition of the matrix A we get that for $i = 2, 3$ we have $-bz_1 + z_3 = z_1 - bz_2 = 0$. Combined with $z_1 + z_2 + z_3 = 1$ this means $z = A_2 := (b, 1, b^2)/(1 + b + b^2)$.

Similarly when $w \in (e_2, e_1)$ respectively $w \in (e_1, e_3)$ we get that z is equal to $A_1 = (1, b^2, b)/(1 + b + b^2)$ and $A_3 = (b^2, b, 1)/(1 + b + b^2)$.

Similarly, if $x(T_k)$ converges, say, to e_3 and simulataneously $z(T_k) \to z$ then we obtain (following the same argument as above) that $\dfrac{\log(x_i(T)) - \log(x_i(0))}{T} \to 0$ for $i = 3$ and therefore

$$\sum_j a_{3j} z_j = 0.$$

This means $z_1 - bz_2 = 0$ (which corresponds to a line segment containing the points A_3 and A_2. So during the very long time interval when $x(T)$ stays near e_3, the average $z(T)$ travels along this segment between A_3 and A_2, so to one of the sides of the triangle $A_1, A_2, .A_3$. $\square$

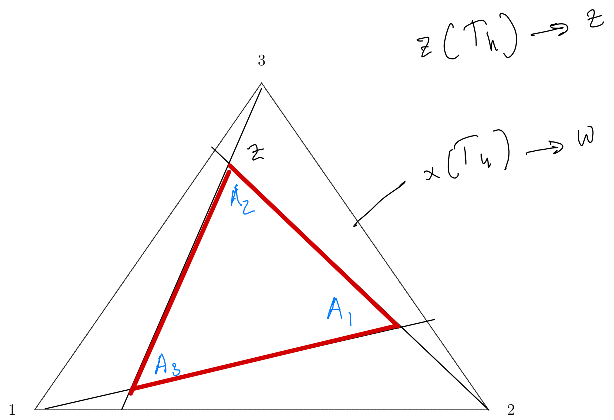


Figure 13: The triangle from Lemma 1.5. As the orbit $x(T_k)$ converges to the boundary of Δ the average $z(T_k)$ converges to the shown triangle. During the very long time interval when $x(T_k)$ converges to one of the corners e_i the average $Z(T_k)$ converges to one of sides of the triangle.

Later on we will consider non-symmetric rock-paper-scissor games, and ask whether these can lead to chaotic dynamics.

Exercise 1.6. 1. Explain why the name Rock-Scissor-Paper is appropriate for the game from example 1.11.

2. Show that A_i, A_{i+1}, e_{i+1} from Lemma 1.5 are collinear.

3. Go through each of the steps in the proof of Lemma 1.5 carefully. For example, give a convincing argument why $\lim_{T\to\infty} \frac{1}{T} \int_0^T x \cdot Ax \, dt \to 0$. (This is a more challenging exercise. Hint: use that there for each corner point there exists a C^1 diffeomorphism which conjugates solutions of this differential equation at that corner point to the solutions of a linear differential equation corresponding to a diagonal matrix.)

1.7 Hypercycle equation and permanence

Is it possible that orbits don't converge to the boundary (we shall call this property *'permanence'*) and also not to a Nash equilibrium in the interior?

Let us consider an example of such a situation. Consider

$$A = \begin{pmatrix} 0 & 0 & 0 & . & . & . & k_1 \\ k_2 & 0 & 0 & . & . & . & 0 \\ 0 & k_3 & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & k_n & 0 \end{pmatrix}$$

This matrix is used to model n populations where population i catalyses the reproduction of the population $i+1(modn)$. This model was created to understand the replication of RNA fragments in the primordial soup. For more on this see Eigen M. and Schuster P., *The Hypercycle* (Springer-Verlag, New York, Berlin, 1979) and https://en.wikipedia.org/wiki/Hypercycle_(chemistry).

To simplify the analysis we will consider the case that $k_i = 1$. So the replicator dynamics is described by

$$\dot{x}_i = x_i(x_{i-1} - \sum_{j=1}^{n} x_j x_{j-1}). \tag{1.14}$$

where we use the cyclic notation, i.e. we take $x_0 = x_n$.

Lemma 1.6. This system has an interior Nash equilibrium which is stable for $n \leq 4$ and unstable (of saddle-type) for $n \geq 5$.

Proof. $E = (1/n)(1, 1, \dots, 1)$ is a Nash equilibrium. An elementary calculation show that the linear part of the system at this point is the matrix

$$\begin{pmatrix} -2/n^2 & -2/n^2 & -2/n^2 & . & . & -2/n^2 & 1/n - 2/n^2 \\ 1/n - 2n^2 & -2/n^2 & -2/n^2 & . & . & . & -2/n^2 \\ -2/n^2 & 1/n - 2/n^2 & -2/n^2 & . & . & . & -2/n^2 \\ . & . & . & . & . & . & . \\ -2/n^2 & -2/n^2 & -2/n^2 & . & . & 1/n - 2/n^2 & -2/n^2 \end{pmatrix}.$$

This is an example of a circulant matrix

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & . & . & . & c_{n-1} \\ c_{n-1} & c_0 & c_1 & . & . & . & c_{n-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ c_1 & c_2 & c_3 & . & . & . & c_0 \end{pmatrix}.$$

It is easy to check that the eigenvalues of such a matric are equal to

$$\gamma_k = \sum_{j=0}^{n-1} c_j \lambda^{jk}, k = 0, \dots, n-1 \tag{1.15}$$

and corresponding eigenvectors

$$(1, \lambda^k, \lambda^{2k}, \dots, \lambda^{(n-1)k}), k = 0, \dots, n-1$$

where $\lambda = e^{2\pi i/n}$. In our setting this leads to $\gamma_0 = -1/n$ and

$$\gamma_k = \sum_{j=0}^{n-1} \frac{-2}{n^2} \lambda^{jk} + \frac{1}{n} \lambda^{(n-1)k} = \frac{\lambda^{-k}}{n}, k = 1, \dots, n-1$$

where we use that the first sum in this expression vanishes. The eigenvalue $-1/n$ with the eigenvector $(1,1,1,\ldots,1)$ (when $k=0$) corresponds to the motion orthogonal to the simplex Δ so is not of interest. When $n=3$, we get $\gamma_k=(1/3)e^{-2k\pi i/3}$, $k=1,2$. The real parts of both these eigenvalues are negative, so the singularity is stable. When $n=4$, we get $\gamma_k=(1/4)e^{-2k\pi i/4}$, $k=1,2,3$ and we see that the eigenvalues γ_1,γ_3 lie on the imaginary axis. Using a Lyapounov function, we will show below that in spite of this the singularity is stable when $n=4$. For $n\geq 5$ there are eigenvalues with positive real part. So in this case the singularity is of saddle type, i.e. both the stable and unstable manifold of the singularity is non-empty. So the singularity is certainly not locally stable.

Let us explain that the boundary of the simplex Δ is repelling (and prove this statement for $n=2,3,4$). Define the Lyapounov function $P(x)=x_1\cdots x_n$. This function is zero on the boundary of Δ and positive in the interior of Δ. Using the chain rule we get that

$$\frac{d}{dt}(\log P)=\sum_{i=1}^n\frac{\dot{x}_i}{x_i}=1-n\sum_{j=1}^n x_jx_{j-1}.$$

For $n=2$ and $n=3$ this function is strictly positive on the interior of Δ except at the Nash equilibrium E. When $n=4$ then

$$\begin{aligned}\frac{\dot{P}}{P}&=\frac{d}{dt}(\log P)=\textstyle\sum_{i=1}^n\frac{\dot{x}_i}{x_i}=1-n\sum_{j=1}^n x_jx_{j-1}=\\&=1-4(x_1x_4+x_2x_1+x_3x_2+x_4x_3)=\\&=1-4(x_1+x_3)(x_2+x_4)=1-4t(1-t)\geq 0\end{aligned}\tag{1.16}$$

where $t=(x_1+x_3)=1-(x_2+x_4)\in[0,1]$. Note that $1-4t(1-t)\geq 0$ and is 0 if and only if $t=1/2$. So

$$\dot{P}/P>0 \text{ on } \operatorname{int}\Delta\setminus W$$

where

$$W:=\{x\in\Delta;(x_1+x_3)=1/2=(x_2+x_4)\}.$$

Claim: this implies that the Nash equilibrium is stable (and in fact attracts all orbits starting in the interior of Δ).

Proof of Claim: Take $x(0)$ in the interior of Δ and suppose by contradiction that $x(t)\not\to E$. By (1.16) and since P has a unique maximum at E, we have then that $t\mapsto P(x(t))$ is increasing while at the same time there exists $\delta>0$ so that $0<P(x(0))\leq P(x(t))\leq P(E)-\delta$ for all $t\geq 0$. Hence $\dot{P}\to 0$ as $t\to\infty$. Using again (1.16) it follows that $x(t)\to W$ as $t\to\infty$. But this implies that for each accumulation point y of $x(t)$ (as $t\to\infty$) is contained in $W=\{x\in\Delta;x_1+x_3=x_2+x_4\}$ Hence the solution $y(t)$ with initial value $y(0)=y$ stays in W for all $t\geq 0$. According to the exercise below, this implies that $y(0)=E$. Hence each accumulation of $x(t)$ is equal to E, which is a contradiction.

Notice that for any arbitrary n

$$\frac{d}{dt}(\log P)=\sum_{i=1}^n\frac{\dot{x}_i}{x_i}=1-n\sum_{j=1}^n x_jx_{j-1}>0$$

whenever $x\in\Delta$ is close to one of the corners e_i of Δ, and so $x(t)$ moves then away from the boundary of Δ. As part of the first project you are asked (based on Hofbauer and Sigmund's book) to show that orbits move away from the boundary when $n\geq 5$. So for $n\geq 5$, the attracting set is neither the boundary of Δ nor the singularity E. $\square$

Remark 1.1. In fact, in a 1991 paper by Hofbauer, Mallet-Paret and Smith it is proved that the ω-limit of any orbit is either an stationary point or a periodic point.

Exercise 1.7.
1. Show that the expression in (1.15) indeed gives the eigenvalues for the matrix C from Lemma (1.6).

2. Show that when $n = 4$ indeed E is the only forward invariant subset of the set W defined in the proof of the previous lemma.
Hint: define the new variable $z = x_1 + x_3 - (x_2 + x_4)$. Then $\dot{z} = (x_3 - x_1)(x_2 - x_4) - z(\sum_{j=1}^n x_j x_{j-1})$. We need to show that if $z(t) = 0$ for all $t \geq 0$ then $z(t) = E$ for all $t \geq 0$. Since $\sum_{j=1}^n x_j x_{j-1} > 0$ for x in the interior of Δ, $z(t) = 0$ for all $t \geq 0$ only holds if $(x_1(t) - x_3(t))(x_2(t) - x_4(t)) = 0$ for all $t \geq 0$. Claim: if this equality holds for all $t \geq 0$ then $x_1(t) = x_2(t) = x_3(t) = x_4(t)$ for all $t \geq 0$ and so $x = E$.

   Proof of claim: If $x_1 - x_3 = 0$ then $\dot{x}_1 - \dot{x}_3 = (x_1x_4 - x_3x_2) - (x_1 - x_3)(\sum_{j=1}^n x_j x_{j-1})$ reduces to $\dot{x}_1 - \dot{x}_3 = x_1(x_4 - x_2)$. Hence $x_1 - x_3 = 0$ and $\dot{x}_1 - \dot{x}_3 = 0$ implies $x_4 - x_2 = 0$. Similarly if $x_2 = x_4$ then $\dot{x}_2 - \dot{x}_4 = x_2x_1 - x_4x_3 - (x_2 - x_4)(\sum_{j=1}^n x_j x_{j-1})$ reduces to $\dot{x}_2 - \dot{x}_4 = x_2(x_1 - x_3)$ and therefore $\dot{x}_2 - \dot{x}_4 = 0$ and $x_2 - x_4 = 0$ implies $x_1 - x_3 = 0$.

1.8 Existence and the number of Nash equilibria

In this section we will show that each game has a Nash equilibrium. There are quite a few proofs of this result. Nash, who was awarded a Nobel prize for introducing the notion of Nash equilibrium invented several proofs of his existence theorem. One of the most well-known proofs is via the Kakutani fixed point theorem, which implies that the multi-valued map

$$\Delta \ni z \mapsto \mathcal{B}R(x)$$

has a fixed point (which by definition is a Nash equilibrium). The two assumptions that need to be checked to see that one can apply the Kakutani fixed point theorem is that (i) the map $\mathcal{B}R$ has a closed graph and (ii) the set $\mathcal{B}R(z)$ is non-empty and convex for each $z \in \Delta$.

In this section we will prove a stronger result, namely that in addition to the existence of a NE, that for "most games" the number of Nash equilibria is odd. To do this, we will prove a result which will assign to each Nash equilibrium an index, and state that the sum of the indices is equal to $(-1)^{n-1}$ where n is the number of dimensions.

To discuss this, we will need to discuss some background on degree theory on the index of a vector field. Several results on this background will not be covered in these lectures.

To start with, let assume that M, N are smooth connected orientable manifolds. If you don't know what a manifold is, then think of for example $M = \mathbb{R}^n$, M is an open ball in $\mathbb{R}^n$, $M = S^1$, $M = S^2$ or the two-dimensional torus $M = T^2$.

Moreover, let $f\colon M \to N$ be a smooth map. We say that y is a *regular value* if $f^{-1}(y) \neq \emptyset$ and if for each $x \in f^{-1}(y)$ the map f is locally smoothly invertible near x, i.e. Df_x is an invertible matrix. In this case, define

$$\text{sign}\, Df_x = \begin{cases} +1 & \text{if } \det(Df_x) > 0 \\ -1 & \text{if } \det(Df_x) < 0 \end{cases}$$

Definition. Let $f\colon M \to N$ be a smooth map and that y is a regular value. Then the degree of f at a point $y \in f(M) \setminus f(\partial M)$ is equal to

$$\deg(f; y) = \sum_{x \in f^{-1}(y)} \text{sign}\, Df_x.$$

Example 1.12. Let $S^1 = \mathbb{R}/\mathbb{Z}$ and $f\colon S^1 \to S^1$ be defined by $x \mapsto nx$. Then deg$(f, y) = n$ for each y. Let M be an open ball in $N := \mathbb{R}^n$ and define $f(x) = -x$. Then deg$(f, y) = (-1)^n$ for each $y \in N = f(M)$.

Assume that M is a compact manifold without boundary (we will only need to consider the case that M is a sphere).

Theorem 1.2. The degree of a map has the following useful properties.

- The degree deg(f, y) of $f\colon M \to N$ is the same for each regular value y of f, see figure in lecture. So this why we speak also of deg(f).
- If $f_t\colon M \to N$ is a family of smooth maps depending smoothly on t, then deg$(f_0) =$ deg(f_1).

Definition. Consider $X\colon \mathbb{R}^n \to \mathbb{R}^n$, and assume x_0 is an isolated zero of X. We will view X as a vector field, so at each point $x \in \mathbb{R}^n$ we have a vector $X(x)$. Take a small sphere S centered at x_0 (i.e. take the boundary of a small ball centered at x_0)) on which X has no zeros, and define the map

$$f\colon S \to S^{n-1}, \text{ by } f(x) = \frac{X(x)}{|X(x)|}.$$

Then the *index* of X at x_0 is defined as

$$\text{ind}(X, x_0) := \text{deg}(f).$$

The same definition applies if X is a vector field on a manifold.

Example 1.13. The index of a saddle point in $\mathbb{R}^2$ is -1, and of a source and a sink is 1.

Lemma 1.7. Assume that X is a vector field, and x_0 an isolated singularity and that its linearisation $A := DX(x_0)$ is non-singular (i.e. invertible). Then ind(X, x_0) is equal to the sign of the determinant of A.

In particular, we have ind$(-X, x_0) = (-1)^n \cdot$ ind(X, x_0) where n is the dimension.

It is easy to check this in dimension two or for linear vector fields. The general case can be seen by deforming the vector field continuously to the linear one, without introducing new singularities.

Example 1.14. Consider the vector field $X(x) = -x$ on $\mathbb{R}^n$. This corresponds to the differential equation $x_i' = -x_i$, $i = 1, \ldots, n$. Then according to the previous theorem, ind$(X, 0) = (-1)^n$. Moreover, the associated map is equal to $f(x) = -x$, and so again deg$(f, 0) = (-1)^n$.

The following remarkable theorem is related to the famous Brouwer fixed point theorem.

Theorem 1.3 (Poincaré-Hopf theorem)**.** Let X be a vector field which is defined on a compact manifold M (you may assume that M is a compact subset of $\mathbb{R}^n$), and assume that if M has a non-empty boundary then for each $x \in \partial M$ one has that $X(x)$ points outwards.

Then

$$\sum_{x, X(x)=0} i_X(x) = \chi(M)$$

where $\chi(M)$ is the Euler characteristic of M.

In this course we won't develop the machinery to compute (or even to formally define) the Euler characteristic of a space. For this you need some homology theory, a subject which is covered in most courses on algebraic homology, and so outside the scope of this course. However, let us give some examples.

Example 1.15. The sphere S^2 in $\mathbb{R}^3$ has Euler characteristic 2. A surface which is made up of a sphere with g handles, has Euler characteristic $2-2g$. So for example the torus has Euler characteristic 0 and the pretzel Euler characteristic -2. In fact, assume that you describe a surface as a convex polyhedron. Then its Euler characteristic $\chi(\text{surface}) = V - E + F$ where V, E, F are the number of vertices, edges and faces. For example, for a cube $V = 8, E = 12, F = 6$ and so $\chi = 2$ while for a tetrahedron, $V = 4, E = 6, F = 4$ and so again $\chi = 2$.

Similarly, an open or closed ball B in $\mathbb{R}^n$ has Euler characteristic $\chi(B) = 1$ whereas the sphere S^n in $\mathbb{R}^{n+1}$ has Euler characteristic $\chi(S^n) = 1 + (-1)^n$.

Example 1.16. The above theorem implies the ***Brouwer's fixed point theorem*** if we assume that the map involved is smooth. This theorem says that any continuous map $f\colon B \to B$ from a ball in $\mathbb{R}^n$ has a fixed point. Let us assume that f is smooth, B is the unit ball and by contradiction assume that f has no fixed point. Then we can define the vector field $X(x)$ defined by $X(x) = x - f(x)$ has no zeros and points along the boundary to the exterior of B. But this contradicts the Poincaré-Hopf theorem as $\chi(B) = 1$.

Example 1.17. The above theorem also implies the so-called hairy ball theorem: If X is a vector field on S^2 then

$$\sum_{x, X(x)=0} i_X(x) = 2.$$

In particular X has at least one zero. The reason this is called the hairy ball theorem is that it implies that a hairy ball has to have places where the "hair sticks up". Note that the above theorem also implies that it is impossible to have a vector field on S^2 with just one saddle point.

Application to game theory

We say that a singularity x_0 of a vector field X is *regular* if the linear part $A = DX(x_0)$ is invertible, and say that a game is *regular* if at each Nash equilibrium $\bar{x}$, the replicator dynamics has a regular singularity (i.e. the linearisation is invertible - so no zero eigenvalue).

Remark 1.2. Assume that x_0 is a regular singularity of the vector field X and let X_λ is a family of vector fields depending differentiably on λ with $X_0 = X$. Then by the implicit function theorem, there exists a function $\lambda \to x_0(\lambda)$ so that $X(x_0(\lambda)) = 0$. (So the singularity moves smoothly as the parameter varies.)

Theorem 1.4. Each $n \times n$ matrix A has at least one Nash equilibrium. Moreover,

1. if A is a regular game, then the number of its Nash equilibria is odd.

2. Consider a Nash equilibrium $\bar{x}$ of the replicator dynamics $\dot{x} = X(x)$ on the boundary of Δ and let $B = DX(\bar{x})$ its linear part. Then any eigenvalue corresponding to any eigenvector of B which is transversal to the boundary of Δ is negative. Hence the stable manifold of $\bar{x}$ points into the interior of Δ, and the unstable manifold of $\bar{x}$ is either empty or fully contained in $\partial\Delta$.

3. Most $n \times n$ matrices are regular.

Proof. Consider the following slight modification of a replicator equation:

$$\dot{x}_i = x_i((Ax)_i - x \cdot Ax - n\epsilon) + \epsilon. \tag{1.17}$$

and let X_ϵ be the vector field defined by the r.h.s. of this expression. Along $\partial\Delta$, the vector field $X_\epsilon(x)$ has no singularities, and points into the simplex Δ. This means that along $\partial\Delta$ the vector field $-X_\epsilon(x)$ points outwards. So, by the Poincaré-Hopf theorem, the sum of the indices of the singularities of $-X_\epsilon$ is equal to 1. Now note that X and $-X$ have the same singularities, and by Lemma 1.7 at each singularity x_0 we have $\mathrm{ind}(-X_\epsilon, x_0) = (-1)^{n-1}\mathrm{ind}(X_\epsilon, x_0)$ because the dimension of Δ is $n-1$. It follows that for each $\epsilon > 0$, the sum of the indices of the singularities of (1.17) is equal to $(-1)^{n-1}$.

This implies that the number of singularities of X_ϵ is odd: let k, l be the number of singularities with index $+1$ resp -1. Suppose by contradiction that $k+l$ is even. Since $k(+1)+l(-1) = (-1)^{n-1}$ we have $k - l = 1 \mod 2$ and $k + l = 0 \mod 2$. This is impossible and therefore we get that $k + l$ is odd.

Let us now show that every singularity of X_0 is a Nash equilibrium (here we use that X_0 is regular). Indeed, for any singularity $p(\epsilon)$ of (1.17) we have

$$(Ap(\epsilon))_i - p(\epsilon) \cdot Ap(\epsilon)) = n\epsilon - \frac{\epsilon}{p_i(\epsilon)}.$$

Note that for any $\delta > 0$, we have that $n\epsilon - \frac{\epsilon}{p_i(\epsilon)} \leq \delta$ for $\epsilon > 0$ sufficiently small. Hence, for any limit point $\bar{p}$ of $\lim_{\epsilon\to 0} p(\epsilon)$ we have that

$$(A\bar{p})_i \leq \bar{p} \cdot A\bar{p}$$

and so $\bar{p}$ is a Nash equilibrium.

Moreover, if all singularities of the replicator system X_0 are regular, then X_0 has finitely many singularities (each one is isolated). Each of these singularities moves smoothly with ϵ and remains a singularity of X_ϵ, i.e. of (1.17). Moreover, the number singularities of (1.17) remains the same for all $\epsilon \geq 0$ small. This proves the first assertion of the lemma.

To prove the 2nd assertion, let us consider a singularity $\bar{x}$ on the boundary, i.e. with $\bar{x}_i = 0$. Because of the form of the equation, the i-row of the linearisation B is of the form $(0\,0\, \dots z_i \dots 0\,0)$ where $z_i = (A\bar{x})_i - \bar{x} \cdot A\bar{x}$ appears on the i-position and the other terms are zero. It follows that any eigenvector with a non-zero i-component has eigenvalue z_i. Since we assumed that the system is regular and it is a Nash equilibrium, we have $z_i < 0$. Hence the 2nd claim holds.

It is not so hard to prove the 3rd assertion, but we will not do this here. □

Example 1.18. In Example 1.8 we had three Nash equilbria: e_1, E and $[e_1, e_3] \cap Z_{1,3}$. The first one is a sink, the 2nd a source, and the final one a saddle, so with index $+1, +1, -1$. The sum of these numbers is equal to $+1 + 1 - 1 = 1 = (-1)^{3-1}$. In several other examples we had a unique NE which was a sink or source in the interior (or a centre) and so there the theorem also holds.

Exercise 1.8. 1. Check the formula from Theorem 1.3 for simple flows on S^2 and on the two-torus T^2. In the case S^2 consider the north-south flow (each point except the north pole flows to the southpole). In the case of T^2 draw a picture of a similar flow (think of a doughnut standing on its side) and the corresponding north-south flow. This flow now has 4 singularities.

2. Give a game which has infinitely many NE.

3. Give a heuristic argument which shows that if a game has only regular singularities, then under the perturbed flow (1.17) the Nash equilibria (of the original flow corresponding to $\epsilon = 0$) on the boundary move into the interior of Δ and the other singularities of the original system move out of Δ.

4. This is a somewhat open-ended question: Discuss why it is a hard problem to find a NE. A starting point is to do an internet search on 'finding Nash Equilibrium is NP hard'.

2 Two player games

So far we looked at a game with one population with different traits, and analysed whether a particular make-up $\hat{x}$ is 'optimal', in the sense that is a Nash or an ESS equilibrium.

A more general situation is when there are two populations which are competing. In this case we have two matrices A, B and assume that the positions of the two populations are determined by two probability vectors x and y.

2.1 Two conventions for the payoff matrices and the existence of NE

There are two different conventions for these matrices. In the **first convention**, the *payoff* and *best-response maps* for the two populations are

$$\begin{array}{ll} P_A(x,y) = x \cdot Ay\,, & \mathcal{B}R_A(y) = \arg\max_{x \in \Delta_A} x \cdot Ay, \\ P_B(x,y) = y \cdot Bx\,, & \mathcal{B}R_B(x) = \arg\max_{y \in \Delta_B} y \cdot Bx. \end{array} \tag{2.1}$$

Here A is be a $n \times m$ matrix and B a $m \times n$ matrix, which means that player A has n strategies and B has m strategies to choose from, and Δ_A, Δ_B are the probability vectors in $\mathbb{R}^n$ respectively $\mathbb{R}^m$. (Often we will assume that $n = m$ and write Δ instead of Δ_A, Δ_B.) We then say that $(\hat{x}, \hat{y})$ is a *Nash equilibrium* iff

$$\hat{x} \in \mathcal{B}R_A(\hat{y}) \text{ and } \hat{y} \in \mathcal{B}R_B(\hat{x}).$$

An equivalent definition is to say that $(\hat{x}, \hat{y})$ is a NE if for all $x \in \Delta_A$ and $y \in \Delta_B$,

$$x \cdot A\hat{y} \le \hat{x} \cdot A\hat{y}\ ,\ \ y \cdot B\hat{x} \le \hat{y} \cdot B\hat{x}$$

If both inequalities are strict if $x \neq \hat{x}$ and $y \neq \hat{y}$ then $(\hat{x}, \hat{y})$ is called a *strict Nash equilibrium*. One could also define $(\hat{x}, \hat{y})$ to be an evolutionary stable equilibrium (ESS) if for all $\epsilon > 0$ and all $x \in \Delta_A \setminus \{\hat{x}\}, y \in \Delta_B \setminus \{\hat{y}\}$,

$$x \cdot A(\epsilon y + (1-\epsilon)\hat{y}) < \hat{x} \cdot A(\epsilon y + (1-\epsilon)\hat{y}),$$

$$y \cdot B(\epsilon x + (1-\epsilon)\hat{x}) < \hat{y} \cdot B(\epsilon x + (1-\epsilon)\hat{x}).$$

In the first convention a pair of matrices is called zero-sum if $A + B^{tr} = 0$.

In fact, there is also **2nd convention** for defining the payoff and best response of the two players, namely as

$$\begin{array}{ll} P_A(x,y) = x \cdot Ay\,, & \mathcal{B}R_A(y) = \arg\max_{x \in \Delta_A} x \cdot Ay \quad \text{and} \\ P_B(x,y) = x \cdot By\,, & \mathcal{B}R_B(x) = \arg\max_{y \in \Delta_B} x \cdot By. \end{array} \tag{2.2}$$

In *this* convention A, B are both $n \times m$ matrices and the definition of NE is as before. $(\hat{x}, \hat{y})$ is called an evolutionary stable equilibrium (ESS) if for all $\epsilon > 0$ and all $x \in \Delta_A \setminus \{\hat{x}\}, y \in \Delta_B \setminus \{\hat{y}\}$,

$$x \cdot A(\epsilon y + (1-\epsilon)\hat{y}) < \hat{x} \cdot A(\epsilon y + (1-\epsilon)\hat{y}),$$

$$(\epsilon x + (1-\epsilon)\hat{x}) \cdot By < (\epsilon x + (1-\epsilon)\hat{x}) \cdot B\hat{y}.$$

In the 2nd convention, a zero-sum game corresponds to $A + B = 0$. The convenience of the 2nd convention becomes clear in the following example:

Example 2.1. Let us consider the situation where both players have two strategies and so the payoff matrices are 2×2: $A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$ and $B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}$. If we use the 2nd convention from (2.2) then one can combine these two matrices using the following notation $\begin{pmatrix} (a_1, b_1) & (a_2, b_2) \\ (a_3, b_3) & (a_4, b_4) \end{pmatrix}$. This corresponds to

$$\left(\begin{array}{l|cc} \text{Payoff's} & \begin{array}{c}\text{Player B}\\ \text{chooses left}\end{array} & \begin{array}{c}\text{Player B}\\ \text{chooses right}\end{array} \\ \hline \text{Player A chooses top} & (a_1, b_1) & (a_2, b_2) \\ \text{Player A chooses bottom} & (a_3, b_3) & (a_4, b_4) \end{array}\right). \tag{2.3}$$

In fact, this 'compact notation' putting the payoff matrices for both players into one box, is only used when referring to the 2nd convention. Note that player A chooses the probability vector x and player B chooses the vector y.

Using the same method as before one can prove:

Theorem 2.1. Each bimatrix game (A, B) has a Nash equilibrium.

Proof. There are quite a few proofs of this result. Several are based on the Brouwer fixed point theorem or on a version of this result. For example, one proof goes along the following lines: Take the map

$$\Psi : \Delta \times \Delta \ni (x, y) \mapsto (\mathcal{B}R_A(y), \mathcal{B}R_B(x)).$$

The righthand side is set-value, so one cannot apply Brouwer's theorem. However, the Kakutani fixed point theorem states thay ot enough that the above map has a closed graph and the right hand side is always non-empty, in order to conclude that there exists $(\hat{x}, \hat{y}) \in (\Psi(x), \Psi(y)$. There is also another proof using index arguments, which gives the parity of the number of NE, similar to the proof given in the previous chapter. $\square$

Remark 2.1. If players can choose between infinitely many actions then the notion of a NE needs clarification and additional assumptions are required to guarantee the existence of a NE.

Exercise 2.1.

1. Compute the NE's for the 2×2 game $\begin{pmatrix} (1, -1) & (0, 0) \\ (0, 0) & (-1, 1) \end{pmatrix}$ where we use the 2nd convention.

2. Let (A, B) be a two-person game and assume we use the 2nd notation. Denote by Δ_A, Δ_B the sets of probability vectors in $\mathbb{R}^n$ resp. $\mathbb{R}^m$. Assume that $(x, y) \in \Delta_A \times \Delta_B$ is in the interior of $\Delta_A \times \Delta_B$. Show that (x, y) is a NE if and only all elements of Ay are equal, and similarly all elements of $x'B$ are equal.

3. Consider the game
$$\left(\begin{array}{c|cc} & \text{i} & \text{ii} \\ \hline \text{i} & (2, 2) & (1, 2) \\ \text{ii} & (2, 1) & (2, 2) \end{array}\right),$$
where we use the 2nd convention (we would otherwise not use the 'compact' notation). Show that (e_1, e_1) and (e_2, e_2) are both NE's, but that (e_1, e_1) is not an ESS while

(e_2, e_2) is an ESS. (Hint: $A\begin{pmatrix} 1-\epsilon \\ \epsilon \end{pmatrix} = \begin{pmatrix} 2-\epsilon \\ 2 \end{pmatrix}$ and $((1-\epsilon)\ \ \epsilon)B = ((2-\epsilon)\ \ 2)$ while $A\begin{pmatrix} \epsilon \\ 1-\epsilon \end{pmatrix} = \begin{pmatrix} 1+\epsilon \\ 2 \end{pmatrix}$ and $(\epsilon\ \ (1-\epsilon))B = ((1+\epsilon)\ \ 2)$.)

2.2 Two player replicator dynamics

If we use the first convention for A, B as in 2.1, the replicator dynamics corresponding to two populations is defined as

$$\begin{aligned} \dot{x}_i &= x_i((Ay)_i - x \cdot Ay) \\ \dot{y}_j &= y_j((Bx)_j - y \cdot Bx). \end{aligned} \tag{2.4}$$

If instead we use the 2nd convention (2.2) for the payoff these equations would become

$$\begin{aligned} \dot{x}_i &= x_i((Ay)_i - x \cdot Ay) \\ \dot{y}_j &= y_j((x^{tr}B)_j - x \cdot By) \end{aligned} \tag{2.5}$$

Some people do not like the aesthetics of the latter expression, and therefore prefer to use the first convention.

Exercise 2.2. 1. As in the previous exercise consider $\begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (-1,1) \end{pmatrix}$. Write down the corresponding replicator equations where we are again using the 2nd convention. Note that x, y are both probability vectors in $\mathbb{R}^2$ so $x_1 = 1 - x_2$ and $y_1 = 1 - y_2$. So the $\Delta \times \Delta$ can be parametrised by (x_1, y_1). Draw the phase diagram.

2.3 Symmetric games

Suppose that $n = m$ and that players A, B choose actions e_i, e_j respectively. Then in the 2nd notation they receive payoffs a_{ij} resp. b_{ij}. Note that if players A, B change role, and swap strategies and payoff matrices, then player A would receive a payoff of b_{ji} and player B a payoff of a_{ji}. We say that the game is *symmetric* if the resulting payoff from such a swap is the same, i.e. if

$$a_{ij} = b_{ji} \text{ and } b_{ij} = a_{ji}$$

i.e.

$$A = B^{tr}$$

Then if one uses the 2nd notation for the replicator equations, the equation (2.5) can be written (in the symmetric case) as

$$\begin{aligned} \dot{x}_i &= x_i((Ay)_i - x \cdot Ay) \\ \dot{y}_j &= y_j((Ax)_j - x \cdot Ay) \end{aligned} \tag{2.6}$$

Exercise 2.3. Show that for a symmetric game, if $x(0) = y(0)$ then $x(t) = y(t)$ for all $t \in \mathbb{R}$. In that sense, in Chapter 1 one can say that the population is playing against itself.

2.4 The 2×2 case

Let us consider the 2x2 case, with payoff matrices $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ and use the *first convention*

$$\begin{aligned} \dot{x}_i &= x_i((Ay)_i - x \cdot Ay) \\ \dot{y}_j &= y_j((Bx)_j - y \cdot Bx). \end{aligned}$$

for the moment. This gives

$$\begin{aligned} \dot{x}_1 &= x_1[a_{11}y_1 + a_{12}y_2 - x_1(a_{11}y_1 + a_{12}y_2) - x_2(a_{21}y_1 + a_{22}y_2)] \\ \dot{y}_1 &= y_1[b_{11}x_1 + b_{12}x_2 - y_1(b_{11}x_1 + b_{12}x_2) - y_2(b_{21}x_1 + b_{22}x_2)] \end{aligned}$$

Using $x_1 + x_2 = 1$ and $y_1 + y_2 = 1$, these formulas reduce to

$$\begin{aligned} \dot{x}_1 &= x_1(1 - x_1)[\alpha_1 - y_1(\alpha_1 + \alpha_2)] \\ \dot{y}_1 &= y_1(1 - y_1)[\beta_1 - x_1(\beta_1 + \beta_2)] \end{aligned} \tag{2.7}$$

where

$$\begin{aligned} \alpha_1 = a_{12} - a_{22}, \quad \alpha_2 = a_{21} - a_{11} \\ \beta_1 = b_{12} - b_{22}, \quad \beta_2 = b_{21} - b_{11}. \end{aligned} \tag{2.8}$$

If, instead, we used the 2nd convention then (2.7) would stay the same except that in (2.8) one would have to exchange the terms b_{12} and b_{21}. However, it may be best to parametrise $\Delta \times \Delta$ by x_2 and y_2, because then $x_2 = 0$ (and $y_2 = 0$) correspond e_1. This means that the the interpretation (2.3) which corresponds to

Payoff's	Player y chooses e_1	Player y chooses e_2
Player x chooses e_1	(a_{11}, b_{11})	(a_{12}, b_{12})
Player x chooses e_2	(a_{21}, b_{22})	(a_{22}, b_{22})

, (2.9)

matches the following figure:

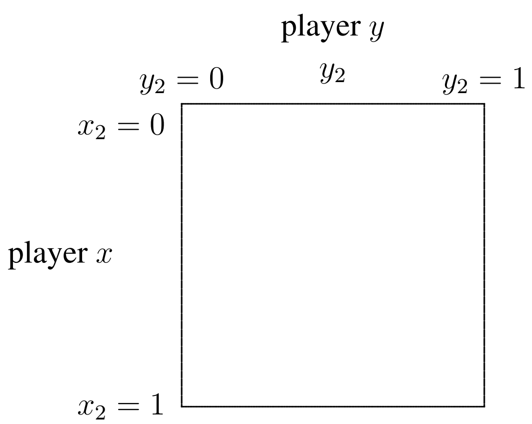


Figure 14: In the 2nd notation, the following parametrisation matches the choice the players make in the decision table (2.9).

This means that, if we use the *2nd convention*, it is more natural to obtain an expression for

$$\begin{aligned} \dot{x}_2 &= x_2((Ay)_2 - x \cdot Ay) \\ \dot{y}_2 &= y_2((x^{tr}B)_2 - x \cdot By) \end{aligned} \tag{2.10}$$

This gives

$$\begin{aligned}\dot{x}_2 &= x_2[a_{21}y_1 + a_{22}y_2 - x_1(a_{11}y_1 + a_{12}y_2) - x_2(a_{21}y_1 + a_{22}y_2)] \\ \dot{y}_2 &= y_2[b_{12}x_1 + b_{22}x_2 - x_1(b_{11}y_1 + b_{12}y_2) - x_2(b_{21}y_1 + b_{22}y_2)]\end{aligned}$$

Using $x_1 + x_2 = 1$ and $y_1 + y_2 = 1$, these formulas reduce to

$$\begin{aligned}\dot{x}_2 &= x_2(1 - x_2)[\alpha_1 - y_2(\alpha_1 + \alpha_2)] \\ \dot{y}_2 &= y_2(1 - y_2)[\beta_1 - x_2(\beta_1 + \beta_2)]\end{aligned} \tag{2.11}$$

where in this setting

$$\begin{aligned}\alpha_1 = a_{21} - a_{11}, \quad \alpha_2 = a_{12} - a_{22} \\ \beta_1 = b_{12} - b_{11}, \quad \beta_2 = b_{21} - b_{22}.\end{aligned} \tag{2.12}$$

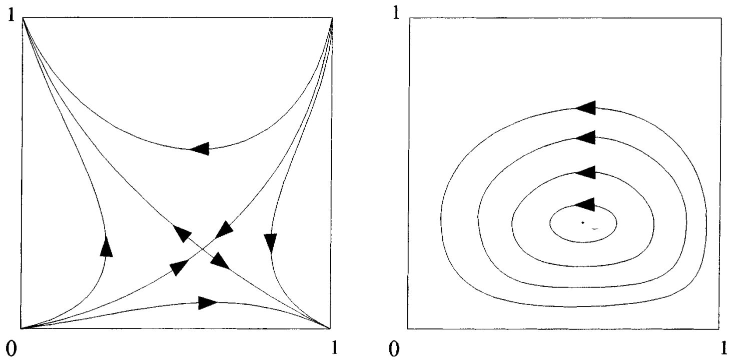


Figure 15: The replicator dynamics of typical 2×2 games.

Using that the r.h.s. of the first equation in (2.7) (resp, (2.11)) has no zeros when $0 < x_1, y_1 < 1$ (resp. when $0 < x_2, y_2 < 1$) when $\alpha_1\alpha_2 < 0$ (and similarly for the 2nd equation), it turns out that there are three possibilities:

Proposition 2.1. There are three possibilities for a 2×2 replicator dynamics system (apart from the degenerate case), namely
(i) $\alpha_1\alpha_2 > 0, \beta_1\beta_2 > 0, \alpha_1\beta_1 > 0$ (coordination games),
(ii) $\alpha_1\alpha_2 < 0$ or $\beta_1\beta_2 < 0$ (dominated strategy, but could be zero-sum),
(iii) $\alpha_1\alpha_2 > 0, \beta_1\beta_2 > 0, \alpha_1\beta_1 < 0$ (zero sum case with interior NE).
The dynamics in case (i) and (iii) is as drawn below.

Exercise 2.4. 1. Explain the terminology used in the previous proposition. Which of the two cases in the above figure corresponds to the battle of the sexes and which is the zero sum case? We already discussed the *prisoner dilemma* game and in the literature also the '*stag hunt*' and '*battle of the sexes*' appear. Zero sum games are often discussed. (Note that often the 2nd rather than the first notation is used when talking about 2×2 games.) What are the NE's and the replicator phase portraits of the following games?

$$\text{stag hunt} : \begin{pmatrix} (4,4) & (1,3) \\ (3,1) & (2,2) \end{pmatrix} \qquad \text{battle of the sexes} : \begin{pmatrix} (3,2) & (0,0) \\ (0,0) & (2,3) \end{pmatrix}$$

$$\text{prisoner dilemma} : \begin{pmatrix} (2,2) & (0,3) \\ (3,0) & (1,1) \end{pmatrix}$$

$$\text{zero-sum}: \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (-1,1) \end{pmatrix} \text{ and } \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (1,-1). \end{pmatrix}$$

2. In the above proposition case (iii) includes zero-sum cases, but it includes also non-zero games. Why is this case still called the 'zero sum case'. (Hint: can there be games with the same phase portrait, but with different matrices?)

3. Why is the interior NE of the replicator differential equation

$$\begin{aligned} \dot{x} &= x(1-x)[\alpha_1 - y(\alpha_1+\alpha_2)] \\ \dot{y} &= y(1-y)[\beta_1 - x(\beta_1+\beta_2)] \end{aligned}$$

either a saddle or orbits are elliptic with orbits cycling around the NE as on the right panel of Figure 15? (Extensive hint: (i) compute the linerarisation of the replicator system at the NE and show that the trace of the linearisation matrix is zero at a Nash equilibrium. (ii) Show that this implies that either both eigenvalues are on the imaginary axis, or one is negative and one is positive - in which case the singularity at this NE is a saddle point. (iii) If both eigenvalues are on the imaginary axis then $\alpha_1\alpha_2 > 0$ and $\beta_1\beta_2 > 0$ and $\alpha_1\beta_1 < 0$. To be definite assume $\alpha_1 > 0$ and $\beta_1 < 0$. Then consider the Lyapounov function

$$P(x,y) = x^{-\beta_1}(1-x)^{-\beta_2} y^{\alpha_1}(1-y)^{\alpha_2}.$$

4. Show that (3) implies that these 2×2 games cannot have an ESS in the interior of the state space (you are allowed use that also in the 2×2 case ESS points are asymptotically stable).

2.5 A 3×3 replicator dynamics systems with chaos (Rock-Paper-Scissors)

A well-known example of a two-player game is

$$A = \begin{pmatrix} \epsilon_x & -1 & 1 \\ 1 & \epsilon_x & -1 \\ -1 & 1 & \epsilon_x \end{pmatrix} B = \begin{pmatrix} \epsilon_y & -1 & 1 \\ 1 & \epsilon_y & -1 \\ -1 & 1 & \epsilon_y \end{pmatrix}$$

Here the first notation is used and we assume that $\epsilon_x, \epsilon_y \in (-1,1)$. Note that this game is zero-sum if $A + B^{tr} = 0$ so when $\epsilon_x + \epsilon_y = 0$. For this game numerical investigations by Sato, Akiyama and coworkers show chaos, see the figures below.

Exercise 2.5. 1. Show that the above game has precisely one NE, namely $(1/3,1/3,1/3),(1/3,1/3,1/3)$. (Hint: make sure you check that there are no NE on the boundaries of the simplices.)

2. The corner points of Δ correspond to R, P, S (paper, rock, scissors). Show that the arrows of the flow along subset $\partial\Delta \times \partial\Delta$ is as in Figure 16.

3. How should you represent the set $\Delta \times \Delta \subset \mathbb{R}^6$ on paper (or on a computer screen)? Show that if you use the linear projection defined by (for example) the matrix

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

then several of the points such as (R, P) and (R, S) will be mapped to the same point. Show that if, instead, you take the linear projection defined by

$$X = \begin{pmatrix} 3.6500 & -1.3500 & 1.3500 & 5.3500 & 1.3500 & 1.4500 \\ 0.4000 & 0.4000 & 4.6000 & 1.9000 & -0.4000 & 4.4000 \end{pmatrix}.$$

then all the nine corner points $(R, R), \ldots, (S, S)$ (which correspond to (e_i, e_j)) map to distinct points in $\mathbb{R}^2$. Check how this relates with Figure 16.

4. Write your own python or matlab code, and produce simulations for the above replicator system. Check whether you obtain similar pictures as the ones shown below. The previous question will be helpful.

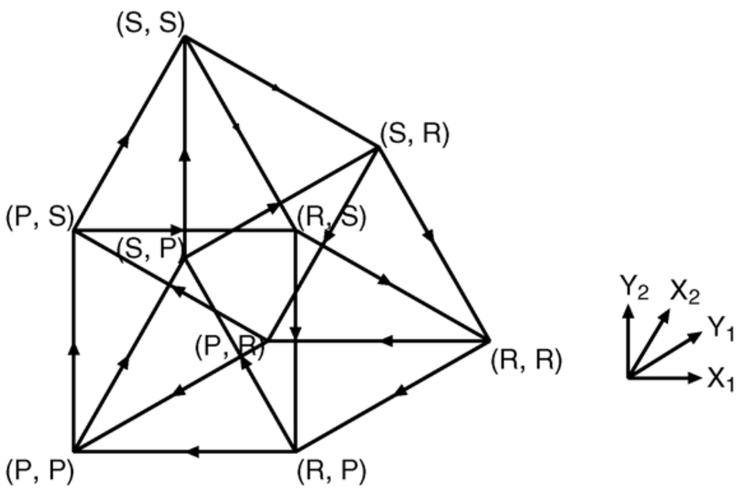


Figure 16: The flow along $\partial\Delta \times \partial\Delta$. This figures come from the paper by Sato et al.

42 *Y. Sato et al. / Physica D 210 (2005) 21–57*

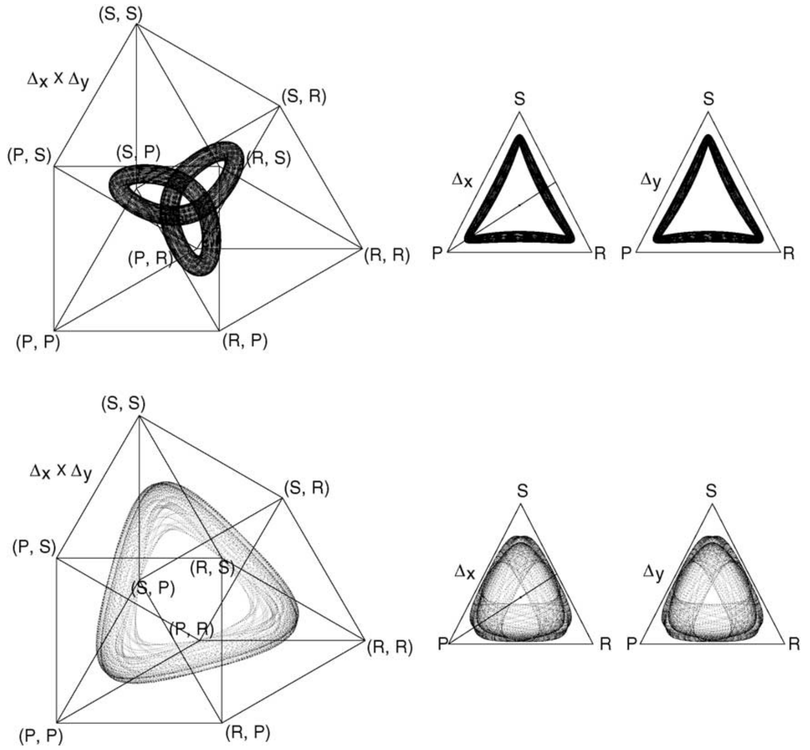


Fig. 12. Quasiperiodic tori: collective dynamics in Δ (left column) and individual dynamics projected onto Δ_X and Δ_Y, respectively (right two columns). Here $\epsilon_X = -\epsilon_Y = 0.0$ and $\alpha_X = \alpha_Y = 0$. The initial condition is (A): $(\mathbf{x}, \mathbf{y}) = (0.26, 0.113333, 0.626667, 0.165, 0.772549, 0.062451)$ for the top and (B): $(\mathbf{x}, \mathbf{y}) = (0.05, 0.35, 0.6, 0.1, 0.2, 0.7)$ for the bottom. The constant of motion (Hamiltonian) is $E = 0.74446808 \equiv E_0$. The Poincaré section used for Fig. 14 is given by $x_1 = x_2$ and $y_1 < y_2$ and is indicated here as the straight diagonal line in agent X's simplex Δ_X.

Figure 17: These figures come from the paper by Sato et al, but you can replicate these figures using the code you can find at the end of the notes.

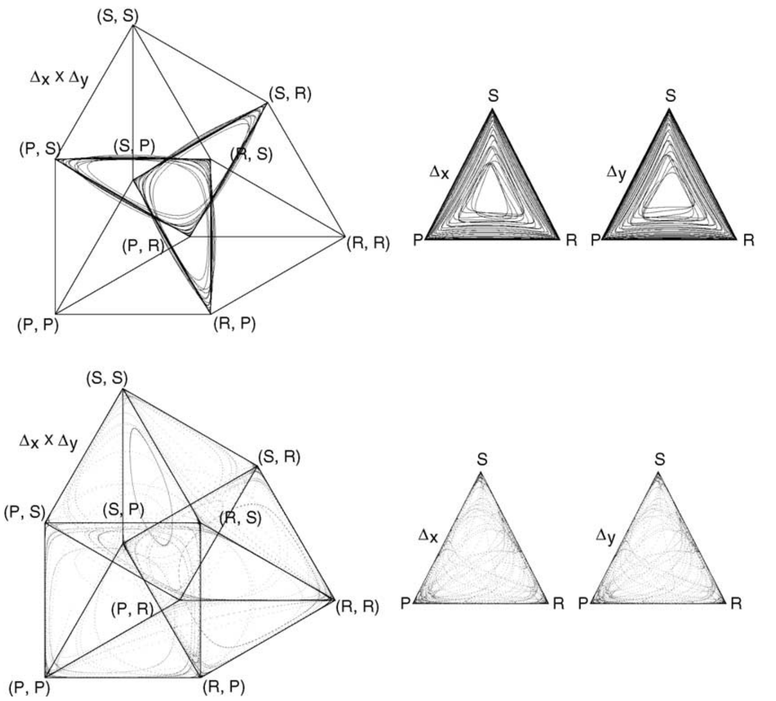


Fig. 15. Heteroclinic cycle with $\epsilon_X = -0.1$ and $\epsilon_Y = 0.05$ (top row). Chaotic transient to a heteroclinic network (bottom row) with $\epsilon_X = 0.1$ and $\epsilon_Y = -0.05$). For both $\alpha_X = \alpha_Y = 0$.

Figure 18: These figures come from the paper by Sato et al, but you can replicate these figures using the code you can find at the end of the notes.

3 Iterated prisoner dilemma (IPD) and the role of reciprocity

In this chapter we will consider the prisoner dilemma game and donation game from the introduction of these notes.[2]

The latter game works as follows: If a player pays c into the scheme the other player receives a benefit of b. So if you cooperate and the other player too then you receive $b-c$, but if you do but the other person not, then you loose $-c$. In particular, the payoff matrices for player I and II are

$$\begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix} \text{ and } \begin{pmatrix} b-c & b \\ -c & 0 \end{pmatrix}$$

where we assume $b > c > 0$, where the strategies are C (cooperate) and D (defect). Here we use the 2nd convention, and therefore the rows are determined by what the first player does and the columns by what the 2nd player does.

A concrete setting might be that you share a house with somebody. To have a tidy house gives you a benefit b, and to tidy it up costs you c. (In this model it is assumed that if you both tidy up, then the costs for each is still c but one can easily modify the pay off matrices to give a cost of $c/2$ if both of you decide to tidy up. What would the payoff matrices look like then?

Of course this is a special case of the general prisoner dilemma game

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix} \text{ and } \begin{pmatrix} R & T \\ S & P \end{pmatrix} \text{ with } T > R > P > S.$$

A frequent choice taken in the prisoner dilemma is

$$\begin{pmatrix} -1,-1 & -3,0 \\ 0,-3 & -2,-2 \end{pmatrix}. \tag{3.1}$$

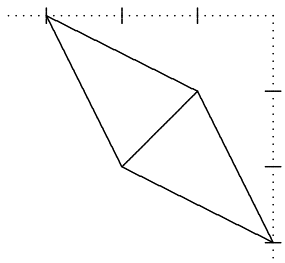


Figure 19: Consider the game (3.1). If the row player chooses actions with probability $(p, 1-p)$ and the column vector with probability $(q, 1-q)$ then the payoff $pq(-1,-1) + p(1-q)(-3,0) + (1-p)q(0,-3) + (1-p)(1-q)(-2,-2)$ of the players lies in the region shown. The payoff $(-1,-1)$ is a Pareto optimum (no player can improve without the other player getting less).

An important feature of both these game is that for the first player the 2nd strategy dominates the first one (i.e. the 2nd row dominates the first one for his matrix) while the 2nd strategy dominates the first one for the 2nd player (i.e. the 2nd column dominates the first one for her matrix). Note that this a symmetric two-player game which will allow us to simplify the discussion.

Of course if this game is repeated exactly 100 times, then by backward induction you can deduce that the best strategy for both players is to never donate (respectively to always defect). But of course for both players to defect 100 times gives for both of them a rather poor payoff of

[2] Watch this ytube clip for a TV programme which employs the prisoner dilemma in a game show.

0 (resp. $100P$) If they always had played the first strategy they would have received $100(b-c)$ (resp. $100R$). So in some sense playing the NE is rather suboptimal. This is also observed by economists, biologists, psychologists etc: often in games which are iterated often players do not choose to play the NE. Why is this? How does altruistic behaviour evolve? In this chapter we will discuss one possible explanation, by consider various scenarios in which you don't know how many times this game is repeated. In such a setting

> the strategy of a player could be to respond to the other player's previous moves.

One such example is the *Tit for Tat strategy*. In the remainder of this chapter an attempt is made to explain why such a Tit for Tat strategy would survive in a competitive environment.

3.1 Repeated games with unknown time length

Let us assume that after each round there is a probability w that the game is repeated at least one more round, where $w \in [0,1]$.

So the probability of the game taking exactly n rounds is $w^{n-1}(1-w)$. This means that the expected duration of the game is

$$1(1-w) + 2w(1-w) + \dots nw^{n-1}(1-w) + \dots = \frac{1}{1-w}.$$

Let us assume that the payoff at round n of the game is equal to A_n and assume that this is bounded. Then the expected total payoff is

$$\sum_{n=1}^{\infty}[A_1 + \dots + A_n]w^{n-1}(1-w).$$

It is easy to see that when $w < 1$ this is equal to the convergent series

$$A(w) := A_1 + wA_2 + w^2A_3 + \dots.$$

Since A_n is bounded, this sum exists and is finite. As the expected duration of the game is $1/(1-w)$, the average payoff per round is therefore

$$(1-w)A(w).$$

Exercise 3.1. Why does it make sense to explore other strategies instead of always to defect if you play a donation game or a prisoner dilemma game for a long time?

3.2 The three strategies AllC, AllD, TFT

Since the game might be played infinitely many times, it makes sense for the players to consider strategies that induce enough trust with the other player so that they will cooperate a lot of the time, because now the issue is not to win but to receive as much pay-off as possible over the duration of the (possible many steps of) the game. For this reason players will invent strategies that take into account the play of the other players.

Let us consider the case where the players consider three strategies: AllC, AllD, TFT. This means Always Cooperate, Always Defect or Tit For Tat (TFT means cooperate if and only if the other player cooperated last time). For simplicity assume (in the TFT strategy) that both players cooperate in the first round.

Then the matrix describing the expected pay-off to the first player is given by

$$\frac{1}{1-w}\begin{pmatrix} b-c & -c & b-c \\ b & 0 & b(1-w) \\ b-c & -c(1-w) & b-c \end{pmatrix} \tag{3.2}$$

where strategies are AllC, AllD, TFT, where as before $b > c > 0$ and $w \in [0, 1)$. Let us check some of the coefficients.

First consider the situation where both players cooperate: then the payoff $A_n = b - c$, $\forall n \geq 1$ and so $A(w) = (b - c)/(1 - w)$.

If both players play TFT then they will keep cooperating, and so the payoff is again $A(w) = (b - c)/(1 - w)$.

If I play TFT and the other player plays AllD, then $A_1 = -c$ and $A_n = 0$, $\forall n \geq 2$, so $A(w) = -c$, which of course is equal to $\frac{1}{1-w}c(1 - w)$.

On the other hand, if I play AllD and the other player TFT, then $A_1 = b$ and from then on $A_n = 0$, so $A(w) = b$.

When we let $w \to 1$ then we get the case where the game is repeated infinitely often.

cm

Exercise 3.2. 1. The matrix (3.2) describes a population which plays a mixture of three strategies. Now suppose that the population also considers the TFTT strategy: only when the other player twice defects will you defect. How would you model this situation?

3.3 The replicator dynamics associated to a repeated game with the AllC, AllD, TFT strategies

Let us consider the replicator dynamics associated to this game, but where we consider the situation where we really have only one population with a given strategy profile, and in which 'individuals' are exploring alternative strategies. So this puts us in the framework of Chapter 1.

For simplicity, let us add to each column of (3.2) a multiple of the vector $\mathbb{1}$. It is easy to see that this does not change the replicator dynamics at all (see Exercise 1.3[2]). Let's apply this operation so that the 2nd row consists of all 0's. This gives

$$\frac{1}{1-w}\begin{pmatrix} -c & -c & bw-c \\ 0 & 0 & 0 \\ -c & -c(1-w) & bw-c \end{pmatrix}.$$

We can also consider the expected average pay-off per round, i.e. multiply the previous matrix by $1 - w$ and consider

$$A = \begin{pmatrix} -c & -c & bw-c \\ 0 & 0 & 0 \\ -c & -c(1-w) & bw-c \end{pmatrix}. \tag{3.3}$$

where as before $b > c > 0$ and $w \in [0, 1)$. The corresponding solutions of the resulting replicator system are then the same, apart from a time reparametrization. Then

$$(Ax)_1 = -c + wbx_3 \text{ , } (Ax)_2 = 0 \text{ and } (Ax)_3 = (Ax)_1 + wcx_2.$$

Note that the best response is always e_2 if $w < c/b$ but that if $w > c/b$ then $\mathcal{BR}(e_1) = e_2$, $\mathcal{BR}(e_2) = e_2$ and $\mathcal{BR}(e_3) =< e_1, e_3 >$.

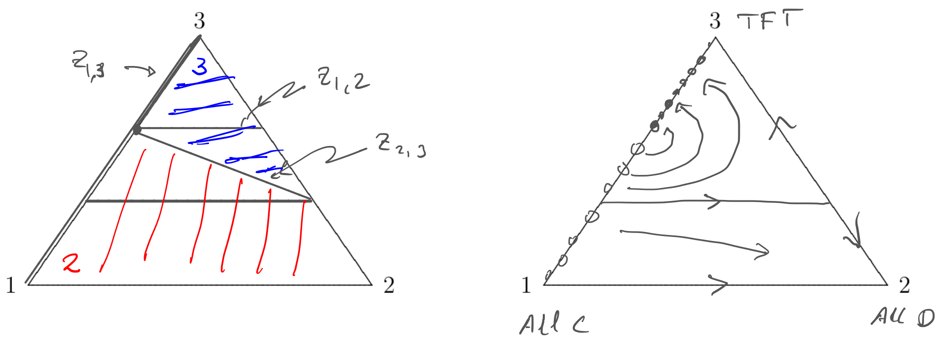


Figure 20: The configurations of the indifference lines corresponding to the repeated the donation game, corresponding to matrix (3.2) (and equivalently to matrix (3.3) for the case that $w > c/b$ and the corresponding phase portrait.

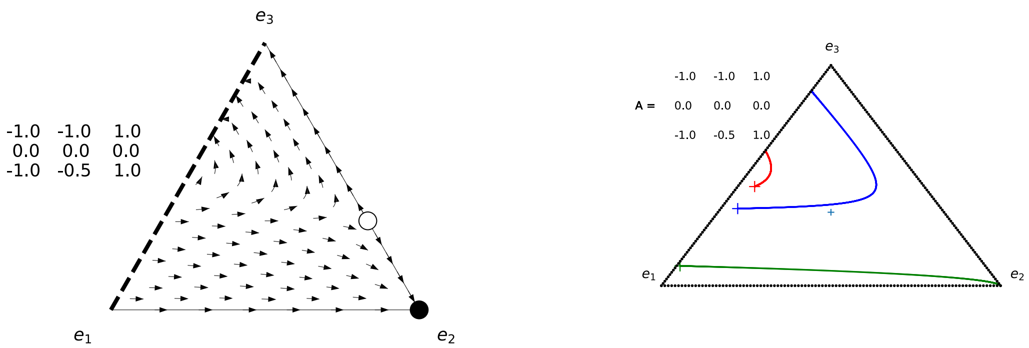


Figure 21: The arrow plot and a computer drawn plot of the flow corresponding to matrix (3.2 where we take $w = 1/2$, $c = 1$ and $b = 4$.

So let us assume that $w > c/b$. Note that $Z_{1,3} = \{x_2 = 0\}$ and that the 3rd row is dominating the 1st one when $x_2 > 0$. Furthermore $Z_{1,2} = \{x_3 = \tilde{x}_3\}$ where $\tilde{x}_3 = c/wb$. Finally $Z_{2,3} = \{cwx_2 + bwx_3 = c\}$ and so is a line connecting $(1 - \tilde{x}_3, 0, \tilde{x}_3)$ to $(0, 1 - \hat{x}_3, \hat{x}_3)$ where $\hat{x}_3 = \dfrac{(1-w)c}{w(b-c)}$ which is $\in (0,1)$ since $w > c/b$.

Using slightly annoying calculations we get

$$x \cdot Ax = (Ax)_3 - x_2 g(x_3), \text{ where } g(x_3) = w(b-c)x_3 - c(1-w). \tag{3.4}$$

Note that $g(\hat{x}_3) = 0$. Because of (3.4) we get that $\dot{x}_3 = x_3[(Ax)_3 - x \cdot Ax] = 0$ along the horizontal line $x_3 = \hat{x}_3$ and so this line is invariant.

Moreover, along this line $g(\hat{x}_3) = 0$ and so if $0 < x_1 < 1$ then we have

$$\dot{x}_2 = x_2[(Ax)_2 - x \cdot Ax] = x_2[0 - (Ax)_3] = -x_2(Ax)_3 > 0$$

because if $x_1 > 0$ then $x_2 + \hat{x}_3 < 1$ and so $(Ax)_3 = -c + cwx_2 + wb\hat{x}_3 < -c + cw(1 - \hat{x}_3) + wb\hat{x}_3 = c(-1 + w) + w(b-c)\hat{x}_3 = 0$.

Along $x_2 = 0$, we have $\dot{x}_2 = 0$ and $(Ax)_3 - x \cdot Ax = x_2 g(x_3) = 0$, so $\dot{x}_3 = \dot{x}_2 = \dot{x}_1 = 0$. It follows that the segment $< e_1, e_3 >$ consists of singularities. The singularities with $x_2 = 0$ and $x_3 \geq \tilde{x}_3 = c/wb$ are attracting (and Nash equilibria) and the the singularities with $x_2 = 0$ and $x_3 < \tilde{x}_3 = c/wb$ are not Nash equilibria.

There are no interior singularities because $(Ax)_3 > (Ax)_1$ when $x_2 > 0$. The best response regions and the solutions of the replicator dynamics are drawn in Figure 20. Note that the AllD solution is not a global attractor.

Exercise 3.3. What are the NE and the ESS for the game corresponding to (3.2) or equivalently (3.3)? This matrix describes a population which plays a mixture of strategies, and discuss why this suggest that it depends on the initial mixture (i.e. the initial condition of the ODE) of the population whether the solution converges to playing always defect (AllD or e_2) or to some mixture of AllC and TFT.

3.4 Random versions of AllC, AllD and TFT

Let us consider a modification of the previous set-up, in which a player makes a probabilistic response to the other players' position. We do this by allocating vectors (f, p, q) and (f', p', q') to the two players. For example, if the first player chooses the TFT strategy from the previous section, then this is described by $(f, p, q) = (1, 1, 0)$. This means that he plays C in the first round ($f = 1$), will definitely reciprocate a C with a C ($p = 1$) but punish a D with a D ($q = 0$).

More formally, $f, f' \in [0, 1]$ gives the probability that the 1st respectively the 2nd player plays C in the first round. Similarly, let p, q be the probability of player I responding in the next round with C when player II plays respectively C, D. So assume that player I cooperates with probability $c(n)$ in round n, then the probability of player II cooperating in round $n+1$ is equal to

$$c'(n+1) = p'c(n) + q'(1 - c(n)) = q' + \rho' c(n)$$

where $\rho' = p' - q'$. The probability of player I cooperating in round $n+2$ is equal to

$$c(n+2) = q + \rho c'(n+1) = \alpha + uc(n)$$

where $\rho = p - q$, $\alpha = q + \rho q'$ and $u = \rho\rho'$. It follows that

$$c(2n+1) = \alpha + u\alpha + \cdots + u^{n-1}\alpha + u^n c(1) = v + u^n(f - v)$$

where f is the probability of player I choosing C in the first round and $v = \alpha/(1-u) = \dfrac{q + \rho q'}{1 - \rho\rho'}$. A similar equation holds for $c(2n+1)$.

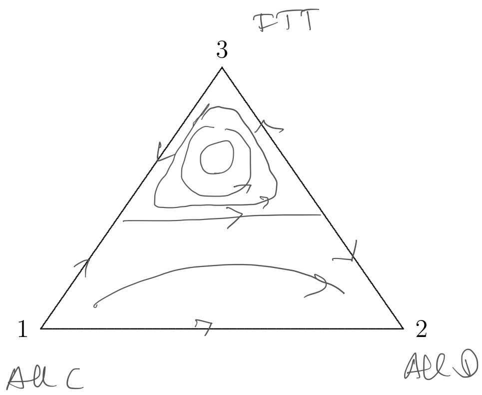


Figure 22: The replicator dynamics corresponding to the matrix (3.5).

In the special case of the donation game (with coefficients $b > c > 0$), and again considering the situation of a probability $1 - w$ after each round of terminating the game, we obtain after some calculations that the average payoff per round of strategy (f, p, q) against (f', p', q') is

$$\frac{-c(e + w\rho e') + b(e' + w\rho' e)}{1 - uw^2}$$

where $e = (1-w)f + wq$, $e' = (1-w)f' + wq'$ (and as before $\rho = p - q$, $\rho' = p' - q'$ and $u = \rho\rho'$). In this way we obtain a fairly complicated 3×3 payoff matrix, which we will not explicitly write down.

Now let us consider one-population replicator dynamics corresponding to this payoff matrix, when there are three possible strategies: $e_1 = (1-\epsilon, 1-\epsilon, 1-\epsilon)$, $e_2 = (k\epsilon, k\epsilon, k\epsilon)$ and $e_3 = (1-\epsilon, 1-\epsilon, k\epsilon)$. Using some simplifications (by adding some multiple of $\mathbb{1}$ to each column to simplify the matrix) we obtain the (normalised) payoff matrix

$$A = \begin{pmatrix} 0 & -1 & \delta\sigma \\ 1 & 0 & -\kappa\sigma \\ \delta & -\kappa & 0 \end{pmatrix} \tag{3.5}$$

where $\delta = w\epsilon$, $\kappa = 1 - w + wk\epsilon$, $\sigma = \dfrac{b\theta - c}{c - c\theta}$ and $\theta = w(1 - (k+1)\epsilon)$.

This gives rise to the replicator dynamics as shown in Figure 22. Note that $\dot{x}_3 = x_3((Az)_3 - x \cdot Ax)$ is zero along some line $x_3 = \hat{x}_3$ Moreover, one can show that there is now cyclic behaviour whenever $x_3(0)$ is large enough, see the exercises below.

Exercise 3.4. 1. Show that $V(x_1, x_2, x_3) = x_1^A x_2^B x_3^C(1 - (1+\sigma)x_3)$ is constant along orbits of the replicator dynamics associated to equation (3.5). Here $A = \kappa/\theta, B = \delta/\theta, C = -1/\theta$. (Hint: use logarithmic derivatives $\dot{x}_i/x_i$.)

3.5 Axelrod tournements: the topic of the 2nd project

In 1980 Axelrod organised a tournement, asking participants to contribute a strategy (a computer programme) which would compete with other strategies all playing a prisoner dilemma game for a large number of iterates. One of the simplest strategies, namely TFT, turned out to do very well. That the TFT did so well was for game theorists quite a surprise because this did not fit in with the classical notion of NE, and because it involved some seemingly 'naive' and 'altruistic' behaviour.

The 2nd project aims to study the question how well different strategies do when battling agains a large number of other strategies. In this project, the question is raised: why does TFT do well and what about an exciting new class of strategies: *zero-deteminant strategies*?

4 No regret learning

In this chapter we will discuss a different class of learning algorithms, namely 'no regret matching' or 'regret minimisation' algorithms. These turn out to learn 'learn' something similar to the (NE), namely the correlated equilibrium (CE) of a bimatrix game. To put this in context, we have the set of Nash equilibria (NE), the correlated equilibria (CE) and the coarse correlated equilibria (CCE). These sets are related as follows:

$$NE \subset CE \subset CCE.$$

In a later chapter, we will introduce the best response dynamics and show that this converges to the CCE set.

4.1 The correlated equilibrium (CE) set

Let us first give the definition of the the CE set. Assume that A has m actions and player B has n actions. We say that the matrix (p_{ij}), $i = 1, \ldots, m$ and $j = 1, \ldots, n$ is a probability distribution if all its entries are ≥ 0 and $\sum_{ij} p_{ij} = 1$. A joint distribution (p_{ij}) is a *correlated equilibrium (CE)* for the bimatrix game (A, B) if

$$\sum_k a_{i'k} p_{ik} \leq \sum_k a_{ik} p_{ik} \quad \text{and} \quad \sum_l b_{lj'} p_{lj} \leq \sum_l b_{lj} p_{lj} \tag{4.1}$$

for all i, i' and j, j'. Note the similarity with the definition of the CCE set defined in Subsection 7.2 where there is a double summation.

This means that if you consider p_{ij} as the proportion of time up to time t that action i, j was chosen, then $t\left(\sum_k a_{ik} p_{ik}\right)$ is the payoff up to time t resulting from action i. The first inequality above means that player A would not have been better off by switching action i to action i'. The 2nd inequality means that the same holds for player B.

Often the notion of CE is motivated by introducing a *trusted intermediator* into the game, who will instruct both players to pick a joint action chosen randomly according to a probability distribution (p_{ij}). This distribution (p_{ij}) is a CE if no player has an incentive to deviate from the intermediator's instructions.

Note that a Nash equilibrium corresponds to the special case where (p_{ij}) is a product distribution, so corresponds to the situation that there are two probability vectors p^*, q^* and that $p_{ij} = p_i^* \cdot q_j^*$, see the exercise below.

It turns out that if both players follow the *no-regret algorithm* which we will discuss in this chapter then the corresponding joint probabilities converge to the CE set. Moreover, if one player plays agains nature (or perhaps against the stock-market) and will use this algorithm, then they will have 'no regret'.

Exercise 4.1. 1. Show that if (p, q) is a NE then the matrix $p_{ij} = p_i q_j$ is in the CE set.

2. Consider the battle of the sexes game: $\begin{pmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{pmatrix}$

where the first action corresponds to watching Tennis and the 2nd one to watching Football. Show that there are two pure NE's for this game, namely (T, T) (with rewards $(2, 1)$) and (F, F) (with rewards $(1, 2)$) and one mixed NE corresponding to probabilities of $p^A = (2/3, 1/3)$ and $p^B = (1/3, 2/3)$ for the two players (with reward $(2/3, 2/3)$). Show that (4.1) corresponds to

$$2p_{21} \leq p_{22},\ \ p_{12} \leq 2p_{11} \quad , \quad 2p_{21} \leq p_{11},\ \ p_{12} \leq 2p_{22}.$$

and therefore to

$$p_{21} \leq (1/2)\min(p_{11}, p_{22}), p_{12} \leq 2\min(p_{11}, p_{22}).$$

Is the set CE finite? In particular, the probability distribution $\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$ is in the CE of this game. Show that the expected payoff of the joint distribution $\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$ is $(3/2, 3/2)$ which obviously outperforms the payoff of the mixed NE for *both players*. Which joint distributions in the CE set have the 'highest' expected payoff (in the sense that is Pareto optimal: if one player would do better, then the other would do worse)? Show that amongst the probability distributions $\begin{pmatrix} 2/9+\epsilon_1 & 4/9+\epsilon_2 \\ 1/9+\epsilon_3 & 2/9+\epsilon_4 \end{pmatrix}$ with $\epsilon_i \approx 0$ the one corresponding to the NE gives the worst pay off for both players. Explain the role of the trusted intermediator to explain the notion of CE in this setting.

3. Consider the

$$\text{game of chicken: } \begin{pmatrix} (6,6) & (2,7) \\ (7,2) & (0,0) \end{pmatrix}.$$

where the first action is to Chicken out and the 2nd action to Dare. What are the NE of this game? Show that the CE inequalities amount to

$$7p_{11} + 0p_{12} \leq 6p_{11} + 2p_{12}, \quad 6p_{21} + 2p_{22} \leq 7p_{21} + 0p_{22}$$

$$7p_{11} + 0p_{21} \leq 6p_{11} + 2p_{21}, \quad 6p_{12} + 2p_{22} \leq 7p_{12} + 0p_{22}$$

Show that the probability distribution $\begin{pmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{pmatrix}$ is in the CE of this game. Explain what your role and that of the role of the trusted intermediator is in this game: what happens to the action (D, D)? Determine the NE of the game, and show that the payoff for $\begin{pmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{pmatrix}$ improves on playing the NE.

4.2 Hart and Mas-Colell's regret matching algorithm

Suppose that the two players have played actions x^i, y^i for time $i = 1, \ldots, t$. Let $\text{SWAP}^i_A(j,k)$ and $\text{SWAP}^i_B(j,k)$ be the payoff at time i which the player would get if they chose action k each time when in fact they played j, assuming that the other player had not changed their action. More precisely, for $i = 1, 2, \ldots, t$, define

$$\text{SWAP}^i_A(j,k) = \begin{cases} e_k \cdot Ay^i & \text{if } x^i = e_j \\ x^i \cdot Ay^i & \text{if } x^i \neq e_j \end{cases}$$

and similarly for SWAP_B. So this gives the payoff A would have received at time i, assuming player B would have done the same, if only they had played k whenever they actually played

j. Note that if $j = k$ then $\text{SWAP}^i_A(j,k) = x^i \cdot Ay^i$.

Then define

$$\text{DIFF}^t_A(j,k) = \frac{1}{t}\left(\sum_{i=1}^{t}[\text{SWAP}^i_A(j,k) - x^i \cdot Ay^i]\right).$$

So this is what player A would have gained (or lost) on average up to time t had they played action k whenever they actually played j. Now define

$$\text{REGRET}^t_A(j,k) = \max(\text{DIFF}^t_A(j,k), 0).$$

Let j^* be the action of player A at time t and let the vector p^{t+1} be defined by

$$\begin{aligned} p_j^{t+1} &= \frac{1}{\mu}\text{REGRET}^t_A(j^*, j) \quad &\text{for all } j \neq j^* \\ p_{j^*}^{t+1} &= 1 - \textstyle\sum_{j\neq j^*} p_j^{t+1} &\text{when } j = j^* \end{aligned} \tag{4.2}$$

Here μ is chosen so large that the above vector is a probability vector. This means that the probability of switching to a different strategy is proportional to their regrets relative to the current strategy. For player B define similarly REGRET^t_B and q^{t+1}.

Theorem 4.1 (Hart and Mas-Colell)**.** Provided we fix μ sufficiently large, if player A follows this algorithm then almost surely $\text{REGRET}^t_A(j,k) \to 0$ as $t \to \infty$.

What this means is that if player A chooses the actions $x^1, \dots, x^t \in \{e_1, \dots, e_n\}$ (where n is the number of actions of player A) according to the probability $p^1, \dots, p^t$ then for each $\epsilon > 0$

$$\mathbb{P}_t(\{(x^1, \dots, x^t) \in \{e_1, \dots, e_n\}^t; \text{REGRET}^t_A(j,k) \geq \epsilon\}$$

goes to zero as $t \to \infty$. Here $\mathbb{P}_t$ is the measure on $\{e_1, \dots, e_n\}^t$ defined by $(p^1, \dots, p^t)$.

Remark 4.1. At first this seems indeed quite strange, so let us make some comments on the algorithm (4.2). As is clear from the exercises in the previous section and this section, this algorithm depends on the actions of the other player (who might be playing random moves). So an action which at time t gives little regret, may give you much more regret later on. So there is no guarantee that a player will stick with a particularly action. To make the previous sentence more specific, suppose you chose action j^* at time t. Then the algorithm defines $p_j^{t+1} = (1/\mu)\text{REGRET}^t_A(j^*, j)$ for $j \neq j^*$ and $p_{j^*}^{t+1} = 1 - \sum_{j \neq j^*} p_j^{t+1}$. Now the above theorem implies that when t is large p_j^{t+1} is small for all $j \neq j^*$. Therefore, by definition, $p_{j^*}^{t+1}$ is close to one. This means that with high probability you will choose at time $t+1$ the same action j^* as at time t. However, if $p_{j'}^{t+1}$ is non-zero for some $j' \neq j^*$, then there is a non-zero probability of choosing this action j' at time $t+1$. If you do, then j' will now replace the role of j^* and most likely you will stick with action j' for a bit. Say, for example, you have two actions. Then you might end up choosing both of them with equal frequency, by repeating the first action many times and then repeating the other action an equal amount of time. This means that you should expect the frequency of the actions not to asymptotically converge to a pure action.

This algorithm even converges to the CE set if both players follow it:

Theorem 4.2 (Hart and Mas-Colell)**.** Provided we fix μ sufficiently large, if both players follow this algorithm then almost surely the resulting frequency of (joint) actions up to time t tends to the CE set as $t \to \infty$.

Foster-Fohra and Fudenberg-Levine have related results.

We will not explain the proofs of these theorems, but prove a result quite similar to Theorem 4.1. To do this we will revisit zero-sum games, vector values payoff functions and the Blackwell approachability theorem.

Exercise 4.2. 1. Let us assume that you are involved in a

$$\text{battle of the sexes game: } \begin{pmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{pmatrix}$$

with actions: watching Football resp. Tennis. If at time i the 2nd player chooses F, i.e. $y^i = F$, then

$$[\text{SWAP}^i_A(j,k) - x^i \cdot Ay^i] = \begin{cases} 0 & \text{if } j = k \\ -2 & \text{if } j = F, k = T \\ 2 & \text{if } j = T, k = F \end{cases}$$

whereas if $y^i = T$ then

$$[\text{SWAP}^i_A(j,k) - x^i \cdot Ay^i] = \begin{cases} 0 & \text{if } j = k \\ 1 & \text{if } j = F, k = T \\ -1 & \text{if } j = T, k = F. \end{cases}$$

So

$$\text{DIFF}^t_A(j,k) = \begin{cases} 0 & \text{if } j = k \\ -2f^t_B(F) + f^t_B(T) & \text{if } j = F, k = T \\ 2f^t_B(F) - f^t_B(T) & \text{if } j = T, k = F. \end{cases}$$

where

$$f^t_B(j) = \#\{1 \le i \le t; y^i = j\}/t \text{ for } j \in \{F, T\}$$

and

$$\text{REGRET}^t_A(j,k) = \max(\text{DIFF}^t_A(j,k), 0)$$

For simplicity write

$$f^t = f^t_B(T) \text{ so that } f^t_B(F) = 1 - f^t.$$

So

$$\text{REGRET}^t_A(j,k) = \begin{cases} 0 & \text{if } j = k \\ \max(3f^t - 2, 0) & \text{if } j = F, k = T \\ \max(2 - 3f^t, 0) & \text{if } j = T, k = F. \end{cases}$$

In Figure 23 these functions are drawn. Show what the intersection $f^t = 2/3$ of these graphs has to do with the matrices of the two players. Explain why $|f^{t+1} - f^t| \le$

$\frac{1}{t}$. Suppose player B chooses action T on the prime number times and F on other times. What would player A do? Alternatively suppose player B picks some sequence $n_{i+1} \geq n_i^2$ and plays F resp. T for times $e^{n_i}, \ldots, e^{n_{i+1}-1}$ when i is even resp. odd. Note that then

$$\frac{e^{n_{i+1}} - e^{n_i}}{e^{n_i}} \to \infty.$$

What would player A do? Is the pay-off matrix of player B relevant for the above discussion?

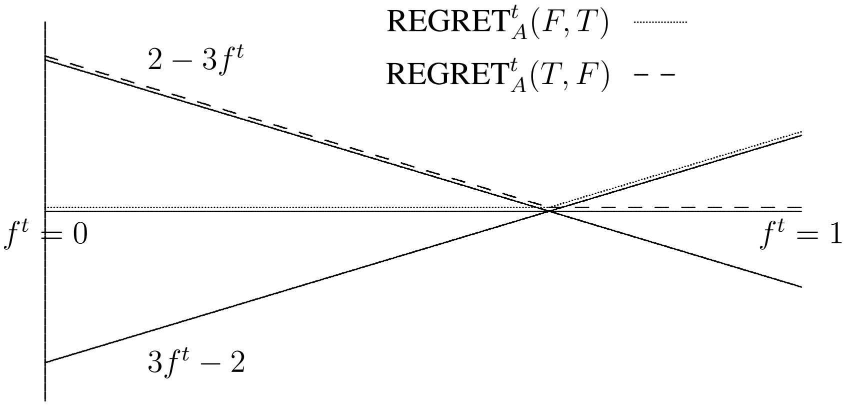


Figure 23: The graphs of REGRET$^t_A(T,F)$ and REGRET$^t_A(F,T)$ as a function of f^t. If $f^t > 2/3$ then you'd choose T to have no regret and if $f^t < 2/3$ then you'd choose F in order to have no regret. If f^t changes over time, you'd adjust your play accordingly.

4.3 Min-max solutions and zero-sum games

Before going into no regret learning it is good to state a well-known fact which is related to zero-sum games.

Theorem 4.3. For any matrix A one has

$$v_A := \max_x \min_y x \cdot Ay = \min_y \max_x x \cdot Ay := v_B. \tag{4.3}$$

If x^*, y^* are so that $\min_y x^* \cdot Ay = v = \max_x x \cdot Ay^*$ then (x^*, y^*) is a NE w.r.t. the two-payer game with matrices A, B where $B = -A$.

Of course the value y for which $\min_y x \cdot Ay$ attains its minimum depends on x, so (4.3) could also be written as $\max_x \min_{y(x)} x \cdot Ay = \min_y \max_{x(y)} x \cdot Ay$.

Remark 4.2. Consider two zero-sum players Alice and Bob with payoff $x \cdot Ay$ and $x \cdot By$ where $B = -A$. Then no matter what Bob does, Alice will get payoff v^A provided she plays

$$x^* \in \arg\max_x \min_y x \cdot Ay.$$

Similarly, Bob will get a payoff of at least $-v^B = -v^A$ proved he plays

$$y^* \in \arg\min_y \max_x x \cdot Ay = \arg\max_y \min_x x \cdot By.$$

$v_A = v_B = x^* \cdot A \cdot y^*$ is called the *value* of the zero-sum game. In view of (4.3) the pair (x^*, y^*) is also called a *minimax value*.

Proof. In the proof of this theorem we will assume the existence of a Nash equilibrium (x^*, y^*) of the game (A, B) where $B = -A$. In fact, as the proof below will show, (4.3) is equivalent to the existence of a Nash equilibrium (x^*, y^*) for zero-sum games.

To prove the $v^A \le v^B$ inequality in (4.3) notice that $\min_y x \cdot Ay \le x \cdot Ay \le \max_x x \cdot Ay$ for all x, y. Hence we have $\min_y x \cdot Ay \le \min_y \max_x x \cdot Ay = v_B$. To prove the opposite inequality, let (x^*, y^*) be a Nash equilibrium of the zero-sum game (A, B) where $B = -A$ (and where we use the 2nd notation). This means $x^* \in BR_A(y^*)$ and $y^* \in BR_B(x^*)$. This is equivalent to the requirement that for all x, y

$$x \cdot Ay^* \le x^* \cdot Ay^* \text{ and } x^* \cdot Ay^* \le x^* \cdot Ay. \tag{4.4}$$

(Note that $B = -A$ and hence the 2nd inequality is $\le$). The previous two inequalities are equivalent to

$$\max_x x \cdot Ay^* = x^* \cdot Ay^* = \min_y x^* \cdot Ay.$$

It follows that we also have the inequality

$$v_B := \min_y \max_x x \cdot Ay \le \max_x x \cdot Ay^* = x^* \cdot Ay^*$$

$$= \min_y x^* \cdot Ay \le \max_x \min_y x \cdot Ay := v_A.$$

This proves the first assertion of the theorem.

In fact, (4.3) implies that there exists a Nash equilibrium. Indeed, take x^*, y^* so that $\min_y x^* \cdot Ay = v = \max_x x \cdot Ay^*$. Let us show that (x^*, y^*) is a NE w.r.t. A, B. Indeed for all x, y,

$$x^* \cdot Ay \ge \min_y x^* \cdot Ay = v \text{ and } x \cdot Ay^* \le \max_x x \cdot Ay^* = v$$

Substituting for x^*, y^* for x, y it follows that $v = x^* \cdot Ay^*$ and

$$x^* \cdot Ay \ge x^* \cdot Ay^* \ge x \cdot Ay^*.$$

and hence the conditions (4.4) for NE are satisfied. □

Exercise 4.3. 1. Let us consider the following zero-sum game

$$\begin{pmatrix} (4, -4) & (-2, 2) \\ (-5, 5) & (6, -6) \end{pmatrix}$$

Let us see how to solve compute the NE in this case from the minimax point of view. Let Alice choose a randomized action $(p, 1-p)$. Then since her payoff is $p'Aq$, this is equal to $4p - 5(1-p) = 9p - 5$ if Bob plays the first action and $-2p + 6(1-p) = 6 - 8p$ if plays chooses the 2nd action. Draw these two lines, and explain why Alice may want to choose $p = 11/17$ corresponding to intersection points of these lines. Similarly, discuss what value Bob will choose for q in his randomized action $(q, 1-q)$.

2. Consider the zero game corresponding to the matrix

$$A = \begin{pmatrix} 4 & 1 & -4 \\ 3 & 2 & 5 \\ 0 & 1 & 7 \end{pmatrix}$$

The coefficient $a_{2,2}$ in this game is called a *saddle-point* of the game, because it is largest in its column and the smallest in its row. Explain why this means that the pure action $(2,2)$ is a NE of the two-player game. Show that it is enough to determine the minima in each row and the maxima in each column, as in

$$A = \begin{pmatrix} -4| & 4 & 1 & -4 \\ 2| & 3 & 2 & 5 \\ 0| & \underline{0} & \underline{1} & \underline{7} \\ & 4 & 2 & 7 \end{pmatrix}$$

and to see whether some value in the new column agrees somewhere with a value in the new row. Note that this approach ONLY works for zero-sum games and their pure NE's. Many zero-sum games do not have a pure NE.

For example, consider the zero sum game corresponding to

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and also w.r.t.

$$A_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Show that that A_1 does not have a pure NE (using the above algorithm) but that A_2 does. Compute the value of these two games.

4.4 Another way of thinking of the minmax theorem

Let A be some function of the form $A\colon \Delta \times \Delta \to \mathbb{R}$. Then

$$\max_q \min_p A(p,q) = \min_p \max_q A(p,q) \tag{4.5}$$

holds iff and only if the following two statements are equivalent for each v:

i. $\exists p \forall q$ s.t. $A(p,q) \leq v$,

ii. $\forall q \exists p$ s.t. $A(p,q) \leq v$.

Note that (4.5) is not the way the minmax statement was formulated in (4.3) if we take $A(p,q) := p \cdot Aq$ where A is a matrix. Indeed, in the previous section it was proved that

$$v_A := \max_x \min_y x \cdot Ay \overset{*}{=} \min_y \max_x x \cdot Ay := v_B$$

and so it first sight it seems that the p,q in (4.5) are the wrong way around. However, if we define $A(p,q) = q \cdot A^{tr} p$ then (4.5) holds. Indeed, $p \cdot Aq = q \cdot A^{tr} q$ and so if we take $p = y$ and $q = x$ in the above displayed formula (and apply this to A^{tr} rather than to A) then

$$\max_q \min_p A(p,q) = \max_q \min_p p \cdot Aq = \max_q \min_p q \cdot A^{tr} p \overset{*}{=} \min_p \max_q q \cdot A^{tr} p =$$

$$= \min_p \max_q p \cdot Aq = \min_p \max_q A(p,q).$$

Exercise 4.4. 1. Show the 'if and only if' statement in the above paragraph.

2. Take $\Delta = [0, 1]$ and let the function $A\colon \Delta \times \Delta \to \mathbb{R}$ be defined by $A(p, q) = pq$. Show that $\max_q \min_p A(p, q) = \min_p \max_q A(p, q) = v = 0$. Show that i) and ii) both hold if $v \geq 0$ and both fail if $v < 0$.

3. Now take $A(p, q) = p + q$ when $p + q \leq 1$ and $A(p, q) = 2 - (p + q)$ otherwise. Show that $\max_q \min_p A(p, q) \neq \min_p \max_q A(p, q)$, so (4.5) fails. (Hint: $\min_p A(p, q) = \min(q, 1 - q)$ and $\max_q A(p, q) = 1$.) This shows that minimax theorem does not hold for arbitrary functions $A(p, q)$. However, some convexity/concavity assumptions on the function A are required (Sion's theorem).

4.5 A vector valued payoff game

The minimax theorem states that if we take $S = \{v : v \geq v^A\}$ where v^A is chosen so that for each y there exists x so that $x \cdot Ay \in S$ then there exists a x^* so that $x^* \cdot Ay \in S$ for all y. So x^* is the silver bullet that deals with all responses!

Suppose player A receives an expected payoff *vector* (rather than a payoff number) depending on the mixed action p and q. So let $A(p, q) \in \mathbb{R}^k$ where we might assume that $A(p, q) = \sum_{i=1}^n \sum_{j=1}^m p_i A_{ij} q_j$ and where $A_{ij} \in \mathbb{R}^k$. Is there an analogue of the minimax theorem?

The Blackwell approachability theorem, which we will next discuss, shows that one can approach such a convex set $\mathcal{C}$, in the sense that of taking longer and longer time averages. For a formal statement see the next subsection.

Example 4.1. Perhaps one way of writing such a map A would be as a matrix with vector entries:

$$A(p, q) = p \cdot \begin{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} & \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 2 \end{pmatrix} & \begin{pmatrix} 1 \\ 5 \end{pmatrix} \end{pmatrix} q$$

which of course is not well-defined. However, we can interpret this as

$$\begin{pmatrix} p \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} q \\ p \cdot \begin{pmatrix} 2 & 3 \\ 2 & 5 \end{pmatrix} q \end{pmatrix} \in \mathbb{R}^2$$

Exercise 4.5. 1. Assume that $A(p, q)\colon \Delta \times \Delta \to \mathbb{R}^k$ is a function and $\mathcal{C}$ is a convex subset of $\mathbb{R}^k$ with the property that for each q there exists p so that $A(p, q) \in \mathcal{C}$. Does this imply that there exists p so that $A(p, q) \in \mathcal{C}$ for all q? (Hint: the answer is: NO if $k \geq 2$!)

2. Let us now assume that in the previous question $k = 1$. So assume that $A(p, q)\colon \Delta \times \Delta \to \mathbb{R}$, where $\Delta = [0, 1]$, and assume A is of the form $A(p, q) = pq$. Assume that $\mathcal{C} \subset \mathbb{R}$ is convex set is so that each q there exists p so that $A(p, q) \in \mathcal{C}$. Show that this implies that there exists p so that $A(p, q) \in \mathcal{C}$ for all q? (Hint: the convex set $\mathcal{C}$ is an interval. What properties does this interval need to have for the assumption to be satisfied. This is related to the previous subsection.)

4.6 Blackwell approachability theorem

Assume that player A decides to play actions according to some probability vectors p^t, $t = 1, \ldots$ and his adversary plays according to q^t, $t = 1, 2, \ldots$. Now assume that the player receives an expected payoff *vector* (rather than a payoff number). Denote this payoff $A(p^t, q^t) \in \mathbb{R}^k$ and let

$$a_t = (1/t) \sum_{t=1}^{t} A(p^t, q^t) \in \mathbb{R}^k.$$

It will be useful to realise that

$$a_t = \frac{t-1}{t} a_{t-1} + \frac{1}{t} A(p^t, q^t).$$

The theorem we will now discuss gives some conditions for payoff vectors to be achieved asymptotically. In this theorem the following scenario is considered:

1. player A chooses an action x^t according to probability p^t
2. player B then subsequently chooses an action according to some probability q^t but without knowing the action x^t.

For simplicity we will assume that the players, in fact, play the mixed action p^t and q^t.

We say that a convex set $\mathcal{C} \subset \mathbb{R}^k$ is *approachable* for the vector payoff A if for each t and all probabilities $\{p^i, q^i\}_{i=1}^{t-1}$ there exists a choice p^t so that for *each* choice of q^t (which player A does not know before choosing p^t), the vectors a_t converge to $\mathcal{C}$ as $t \to \infty$ (in the Euclidean norm). Blackwell's Approachability Theorem gives a necessary and sufficient condition for $\mathcal{C} \subset \mathbb{R}^k$ to be approachable. In the setting of this theorem it will turn out that p^t only depends on $\mathcal{C}$, a^{t-1} and $A(p^{t-1}, q^{t-1})$.

Here we will always assume that $A(p, q)$ can be written as

$$A(p, q) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i A_{ij} q_j$$

but where A_{ij} is a vector.

Theorem 4.4 (Blackwall's Approchability)**.** For any closed convex set $\mathcal{C}$ the following are equivalent.

1. $\mathcal{C}$ is approachable for the vector payoff A;
2. for each q there exists p so that $A(p, q) \in \mathcal{C}$;
3. every half space containing $\mathcal{C}$ is approachable.

Proof. (2) $\implies$ (3) Consider a half-space $H = \{a \in \mathbb{R}^k; n \cdot a \leq v\}$ which contains $\mathcal{C}$ where n is the normal vector to the half-plane H.

$$\forall q \exists p \text{ with } A(p, q) \in \mathcal{C} \implies \forall q \exists p \text{ with } n \cdot A(p, q) \leq v \implies$$

$$\exists p \forall q \text{ with } n \cdot A(p, q) \leq v \implies H \text{ is approachable}$$

Here the exchange of $\forall q \exists p$ to $\exists p \forall q$ follows since the minmax theorem holds for $(p, q) \to n \cdot A(p, q)$ (because $n \cdot A(p, q) = p \cdot \tilde{A} q$ for some matrix $\tilde{A}$) and in the conclusion one chooses the $p^t = p$ where p is from the last line.

(3) $\implies$ (2) Since *each* half-space $H \supset \mathcal{C}$ is approachable, there exists for each such half-space H and for each q some p with $A(p, q) \in H$ (to see this, take $q^t = q$ for all t). Since this holds for each such half-space we also have $\forall q \, \exists p$ with $A(p, q) \in \mathcal{C}$.

(1) $\implies$ (3) trivially follows from $\mathcal{C} \subset H$.

(3) $\implies$ (1) is the most interesting part of the proof. Let $\pi(a_t)$ be the closest point in $\mathcal{C}$ to a_t, let $n_t = a_t - \pi(a_t)$ and let $v_t = \pi(a_t) \cdot n_t$. Then let H_t be the half-space containing $\mathcal{C}$ through $\pi(a_t)$ orthogonal to n_t. That is, $H_t = \{a; a \cdot n_t \leq v_t\}$. Draw a picture!

Since H_t is a half-plane and since in principle player B could take q^t for all t,

$$H_t \text{ is approachable} \implies \forall q \, \exists p \text{ so that } n_t \cdot A(p, q) \leq v_t. \tag{4.6}$$

Note that $n_t \cdot A(p, q)$ is of the form $p \cdot A^t q$ where A^t is a matrix which depends on on n_t. Since the minmax theorem holds for A^t, as we saw in Section 4.4, we can exchange the quantifiers in (4.6) and so we get $\exists p \, \forall q$ so that $n_t \cdot A(p, q) \leq v_t$. Let $p^t = p$ where p is this choice. So

$$n_t \cdot A(p^t, q^t) \leq v_t = \pi(a_t) \cdot n_t \quad \text{and therefore} \quad n_t \cdot (A(p^t, q^t) - \pi(a_t)) \leq 0. \tag{4.7}$$

Let us now show that there exists K so that $||a_t - \pi(a_t)|| \leq K/\sqrt{t}$ for all t and so $a_t \to \mathcal{C}$ as $t \to \infty$. To see this, take $|| \cdot ||$ to be the Euclidean norm. Then

$$d(a_{t+1}, \mathcal{C})^2 \leq ||a_{t+1} - \pi(a_t)||^2 \leq ||\frac{t}{t+1} a_t + \frac{1}{t+1} A(p^{t+1}, q^{t+1}) - \pi(a_t)||^2.$$

Using $n_t = a_t - \pi(a_t)$ the last expression is equal to

$$\begin{aligned}
&= ||\frac{t}{t+1} n_t + \frac{1}{t+1} \left(A(p^{t+1}, q^{t+1}) - \pi(a_t)\right) ||^2 \leq \\
&= (\frac{t}{t+1})^2 ||n_t||^2 + (\frac{1}{t+1})^2 ||A(p^{t+1}, q^{t+1}) - \pi(a_t))||^2 + \\
&\quad + 2\frac{t}{(t+1)^2} \left(n_t \cdot (A(p^{t+1}, q^{t+1}) - \pi(a_t))\right) \\
&\leq (\frac{t}{t+1})^2 ||n_t||^2 + (\frac{1}{t+1})^2 ||A(p^{t+1}, q^{t+1}) - \pi(a_t))||^2.
\end{aligned}$$

where we first used the equality $||u + v||^2 = ||u||^2 + ||v||^2 + 2u \cdot v$ where $\cdot$ stands for the usual inner product and subsequently the inequality (4.7). Since $n_t = d(a_t, \mathcal{C})$ this gives

$$(t+1)^2 d(a_{t+1}, \mathcal{C})^2 \leq t^2 d(a_t, \mathcal{C})^2 + ||A(p^{t+1}, q^{t+1}) - \pi(a_t))||^2.$$

To simplify the proof, let us assume that the set $\mathcal{C}$ is a compact. Since the values of A are bounded, this gives that the 2nd term in the sum is $\leq \hat{K}$. Using a telescopic sum we get

$$(t+1)^2 d(a_{t+1}, \mathcal{C})^2 \leq d(a_1, \mathcal{C})^2 + \hat{K}(t+1)$$

and so there exists K so that

$$d(a_{t+1}, \mathcal{C}) \leq K/\sqrt{t+1}$$

for all $t = 0, 1, 2, \ldots$. $\square$

Exercise 4.6. 1. Draw a diagram which clarifies the previous proof, and shows how the 'algorithm' for choosing p^t suggested in step (3) of the proof works in pseudo-code.

2. Assume that
$$a_t = (1/t)\sum_{t=1}^{t} A(p^t, q^t) \in \mathcal{C}.$$
Explain why it is not possible for player A to guarantee that $a_{t+1} \in \mathcal{C}$. (Hint: use Exercise 7.5.1.)

4.7 Regret minimisation

Let us give an application of the previous Blackwell approachability theorem. Take a real valued payoff $A(p,q)$ of the form $p \cdot Aq$ and consider the vector valued $\hat{A}(p,q) \in \mathbb{R}^n$ with components $A(e_i,q) - A(p,q)$, $i = 1,\dots,n$. So this is the gain or loss if player A would choose the mixed strategy p instead of strategy i.

Now consider the convex region $\mathcal{C} = \{a; a_i \le 0\,\forall i\}$. For each q there exists p so that each of the components of $\hat{A}(p,q) \le 0$: choose $p = e_{i^*}$ where $i^* = \arg\max_i A(i,q)$. It follows that the 2nd condition of Blackwell's approachability theorem is satisfied for the set $\mathcal{C}$. In particular this set is approachable for the payoff $\hat{A}$, and so given $p^1, q^1, \dots, p^{t-1}, q^{t-1}$ there exists a strategy p^t so that for whatever q^t is one has
$$\left(\frac{1}{t}\sum_{s=1}^{t}[A(e_i,q^s) - A(p^s,q^s)]\right)_{i=1}^{n} = (1/t)\sum_{s=1}^{t}\hat{A}(p^s,q^s) \to \mathcal{C} \text{ as } t \to \infty.$$
Hence, for each i,
$$\limsup_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t}[A(e_i,q^s) - A(p^s,q^s)] \le 0.$$
This means that the regret tends to zero (remember that the regret was the positive part of the previous expression).

To complete the proof we would need to do two things

1. To show that one can also show that swapping one particular action for another one would not lead to regret. In the above proof we only considered the case that one always would play action e_i.

2. Extend the argument to show that playing pure actions (which one usually is required to do) according to the mixed probability vectors still would not lead to regret. This step is based on the fact that mixed strategies gives the expected payoff when picking actions according to this mixed strategy.

We will not elaborate on these additional steps in the proof, because they do not give much additional insight and are a little tedious.

Exercise 4.7. 1. Explain why the above strategy does not work if player B can choose q^t after seeing what player A has done at time t.

5 Reinforcement learning

Reinforcement learning (RL) is a very widely used approach to learning. The premise in RL is that you choose an action which in the past turned out to be quite rewarding more often in the future (exploitation), but at the same time you do not rule out choosing actions which were less successful from to time in order not to miss opportunities (exploration). The latter avoids that you get stuck playing a 'local maximum'.

This chapter will describe various models for reinforcement learning, which focusses on actions which gave good payoff's in the past. As we will see, many of these reinforcement learnings have solutions which closely mimic those of replicator systems and related systems.

5.1 Set-up of reinforcement learning

1. at each time period t, each of the two players chooses an action $x(t)$ resp. $y(t)$. Here $x(t), y(t)$ will be probability unit vectors e_i, e_j. For simplicity, often we write x^t, y^t instead of $x(t), y(t)$. In fact, the other player could be 'nature' or a player which has an unknown way of choosing strategies.

2. the payoff (or reward) for player A is given by a function $u^t = u(x^t, y^t) \in \mathbb{R}$ which for pure actions can be written in the form $x^t \cdot Ay^t$ where A is a matrix. In the current chapter we will assume that the *payoff is always strictly positive*: there exists $C_0 > 0$ so that $A_{ij} > C_0$ for all i, j and so $u^t \geq C_0$ for all $t \geq 1$.

3. Define a variable $\theta_i^t \geq 0$ which describes the *propensity* of player A to play action i (i.e. to choose $x^t = e_i$) at time t. The variable θ_i^t is updated in some manner according to how "good playing x" has been. Let $\theta^t = (\theta_1^t, \ldots, \theta_n^t)$. At time t, the probability that A plays action i is determined by θ_i^t. For example one can choose actions according to the probability vector

$$p^t = \frac{\theta^t}{|\theta^t|_1} \tag{5.1}$$

where $|z|_1 := \sum_i |z_i|$ when $z \in \mathbb{R}^n$. Some other ways of choosing actions are discussed in Subsection 5.5.

4. Several updating rules have been proposed for the propensity θ^t. Here we will always assume that $|\theta^1| = C > 0$ and indeed that all coordinates of θ^1 are positive.

5. As mentioned, in reinforcement learning the action $x^t \in \{e_1, \ldots, e_n\}$ are chosen randomly so that $x^t = e_i$ with probability p_i^t.

   One way of visualising (and implementing) such a random variable is to partition the interval $[0, 1]$ into n intervals $I_1^t, \ldots, I_n$ where $I_1^t = [0, p_1^t]$ and $I_i^t = [p_1^t + \cdots + p_{i-1}^t, p_1^t + \cdots + p_i^t]$ for $i > 1$. Then choose r uniformly in $[0, 1]$ and choose $x^t = e_i$ if $r \in I_i^t$.

Three well-known update rules for the propensity function:

1. **Cross-learning**, named after Cross (1973):

$$\theta^{t+1} = (1 - \vartheta u^t)\theta^t + \vartheta u^t x^t, \quad t \geq 1$$

where we assume in this model that $u^t \in (0, 1)$, $\vartheta \in (0, 1]$ and $|\theta^1|_1 = 1$ (and that each component of the vector θ^1 is strictly positive). So in this model, $|\theta^t|_1 = 1$ for all $t \geq 1$.

2. **Erev-Roth Cumulative payoff matching (CPM)** (dating back to 1995) is:

$$\theta^{t+1} = \theta^t + u^t x^t, \quad t \geq 1.$$

Here the initial vector θ^1 is chosen so that $|\theta^1|_1 = C$ and so that each component of θ^1 is strictly positive, see below.

3. The **Arthur model** (dating back to 1993), is closely related to the previous one:

$$\theta^{t+1} = (\theta^t + u^t x^t)\frac{C(t+1)}{Ct + u^t}, \quad t \geq 1.$$

Here $C > 0$ is chosen fixed throughout, and θ^1 is chosen so that $|\theta^1|_1 = C$ and so that each component of θ^1 is strictly positive.

All models have in common that they reinforce playing a particular action depending on the payoff it resulted in previously. Note that player A does not need to observe the actions of player B to determine θ^t, only their own utility pay-off. The vector θ^t can be viewed as some kind of 'score-card' on how well the various actions have done in the past.

In the first part of this chapter, we will focus on the above updating rules where two players use in their interactions and where we shall concentrate on models (2) and (3), which in some sense are quite similar: the updating rule for the latter is just a rescaled version of the former one.

In the next examples we will consider the situation where a player interacts with 'nature'.

Example 5.1. Let us consider in this exercise the situation of a player who interacts with 'nature'. For example, suppose a doctor can prescribe a new medication to a patient, but it is not clear whether this will also have good outcomes. One approach is to do a blind sample, giving some patients a Placebo and others the new Medication, and then to compare the statistics. Another one is to use reinforcement learning. Every time one medication 'worked' you increase its propensity, and start using the more successful medication more often.

Exercise 5.1. 1. Assume that we are in the setting of Example 5.1 and the payoff for M is 10 and P is 5. What would be the probability of giving M be in the limit?

5.2 The Arthur model, in the 2×2 setting

In this subsection we will assume that both players learn according to the model (3) and follow the analysis from Posch (1997). Note that in this model $|\theta^t|_1 = tC$ for all $t \geq 1$. Indeed, $|\theta^1| = C$ by assumption. Assume by induction $|\theta^t|_1 = tC$ then because (by assumption) $u^t > 0$,

$$|\theta^{t+1}|_1 = |\theta^t + u^t x^t|_1 \frac{C(t+1)}{Ct + u^t} = C(t+1).$$

Since $p^t = \theta^t / |\theta^t| = \theta^t/(Ct)$,

$$\begin{aligned} p^{t+1} \quad &:= \frac{\theta^{t+1}}{|\theta^{t+1}|_1} = \frac{\theta^t + u^t x^t}{Ct + u^t} = \frac{Ctp^t + u^t x^t}{Ct + u^t} \\ &= p^t + \frac{u^t}{Ct + u^t}(x^t - p^t) = p^t + \frac{u^t}{Ct}(x^t - p^t) + \frac{1}{Ct}\epsilon^t \end{aligned} \tag{5.2}$$

where $\epsilon^t = O(1/t)$. This is because $|u^t| \le A := \max_{ij} |A_{ij}|$ and therefore

$$\frac{1}{Ct+u^t} = \frac{1}{Ct} - \frac{u^t}{(C+u^t/t)(Ct^2)} \text{ and so } |\frac{u^t}{(C+u^t/t)}| \le \frac{A}{C-A/t} = O(1)$$

for t large. We will also assume that the 2nd player uses the corresponding updating rule for their probability vector q^t.

Note that the actions $x^1, \dots, x^{t-1}$ and $y^1, \dots, y^{t-1}$ determine $u^1, \dots, u^{t-1}$ and therefore $\theta^1, \dots, \theta^t$ and $p^1, \dots, p^t$. The first player chooses action i with probability equal to the i-th coordinate p_i^t of p^t. So $x^t = e_i$ with probability p_i^t. Finally, the payoff u^t is then determined by the action y^t of the 2nd player together with x^t. If we define $f(p^t, q^t)$ to be the conditional expectation

$$f(p^t, q^t) = \mathbb{E}(u^t(x^t - p^t)|\{(x^1, y^1), \dots, (x^{t-1}, y^{t-1})\}),$$

then we can write

$$u^t(x^t - p^t) = f(p^t, q^t) + \mu(p^t, q^t)$$

where $\mu(p^t, q^t)$ is a variable with the property that

$$\mathbb{E}(\mu(p^t, q^t)|\{(x^1, y^1), \dots, (x^{t-1}, y^{t-1})\}) = 0.$$

Hence one can write the previous recurrence equation (5.2) as

$$p^{t+1} = p^t + \frac{1}{Ct}\left[f(p^t, q^t) + \mu(p^t, q^t) + \epsilon^t\right] \tag{5.3}$$

where, as before,

$$\mathbb{E}(\mu(p^t, q^t)|\{(x^1, y^1), \dots, (x^{t-1}, y^{t-1})\}) = 0 \text{ and } \epsilon^t = O(1/t).$$

Note that $\mu(p^t, q^t)$ depends on $x^1, \dots, x^t$ and $y^1, \dots, y^t$.

Writing Equation (5.2) in the form of Equation (5.3) has one draw-back: this equation in principle allows the vector p^t to be no longer a probability vector, whereas in the original expression (5.2) it is clear that that p^t remains a probability vector (and so in particular, each of its components is between 0 and 1). On the other hand, (5.3) makes the connection with a differential equation more clear, as we will see.

5.2.1 A two player version of this with two actions

Now do the same for the other player. Then the action y^t player B chooses depends on the corresponding probability q^t, and so we obtain the discrete time stochastic process:

$$\begin{aligned} p^{t+1} &= p^t + \tfrac{1}{Ct}\left[f(p^t, q^t) + \mu(p^t, q^t) + \epsilon^t\right] \\ q^{t+1} &= q^t + \tfrac{1}{Ct}\left[g(p^t, q^t) + \zeta(p^t, q^t) + \epsilon^t\right]. \end{aligned} \tag{5.4}$$

5.2.2 Stochastic approximation of an ODE

Note that when $\mu = \zeta = \epsilon^t = 0$, (5.4) is the Euler approximation with decreasing time steps of the differential equation

$$\begin{aligned} \dot{p} &= f(p^t, q^t) \\ \dot{q} &= g(p^t, q^t). \end{aligned} \tag{5.5}$$

Since $\mathbb{E}(\mu(p^t, q^t)|\{(x^1, y^1), \ldots, (x^{t-1}, y^{t-1})\}) = 0$ (and similarly for ζ) and $\epsilon^t = O(1/t)$, it is reasonable to expect that the solutions of (5.4) after T steps should should be closely related to those of (5.5) after time $\sum_{t\geq 1}^{T} \frac{1}{(tC)} \approx \frac{1}{C}\log(T) \to \infty$.

This connection between the discrete sequence of random variables (5.4) and the ODE (5.5) is described rigorously in for example Benaim (1999), where one of the main results is the following:

Theorem 5.1. Almost all realisations (p^t, q^t), $t = 1, 2, \ldots$ of (5.4) tend asymptotically to a 'internally chain recurrent set' of the differential equation (5.5).

Exercise 5.2. Give examples of 'internally chain recurrent sets' for the two differential equations drawn in Figure 24.

Notice that the above theorem does not claim that almost all realisations tend to attractors of the differential equations. In Proposition 5.1 it is shown that this indeed need not be the case.

5.2.3 Calculating f and g in the 2×2 case

For simplicity assume that each of the players has two actions and that the payoff matrices are A and B are equal to $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ where we use the 1st convention and derive the deterministic part of the above differential equation. Let us temporarily write $\underline{p}^t, \underline{q}^t$ for the vectors and p^t, q^t for its first components, so write $\underline{p}^t = (p^t, 1-p^t)$ and $\underline{q}^t = (q^t, 1-q^t)$. We will also temporarily write

$$\underline{f}(\underline{p}^t, \underline{q}^t) = \mathbb{E}(u^t(\underline{x}^t - \underline{p}^t)|\{(\underline{x}^1, \underline{y}^1), \ldots, (\underline{x}^{t-1}, \underline{y}^{t-1})\})$$

for the vector and $f(\underline{p}^t, \underline{q}^t)$ for its first component. If $\underline{x}^t = \underline{e}_i$ then $u^t(\underline{x}^t - \underline{p}^t) = (\underline{e}_i \cdot A\underline{q}^t)(\underline{e}_i - \underline{p}^t)$. The probability of this occurring is $\underline{p}_i^t$, where $\underline{p}^t = (\underline{p}_1^t, \underline{p}_2^t) = (p^t, 1-p^t)$. Hence the first component of

$$\mathbb{E}(u^t(\underline{x}^t - \underline{p}^t)|\{(\underline{x}^1, \underline{y}^1), \ldots, (\underline{x}^{t-1}, \underline{y}^{t-1})\}) = \sum_i \underline{p}_i^t((\underline{e}_i \cdot A\underline{q}^t)(\underline{e}_i - \underline{p}^t))$$

is equal to

$$\begin{aligned} f(\underline{p}^t, \underline{q}^t) &= p^t(a_{11}q^t + a_{12}(1-q^t))(1-p^t) + \\ &\quad (1-p^t)(a_{21}q^t + a_{22}(1-q^t)(0-p^t) \\ &= p^t(1-p^t)\left((a_{12}-a_{22}) - q^t((a_{21}-a_{11}) + (a_{12}-a_{22})\right)) \end{aligned}$$

and similarly for g. That is,

$$\begin{aligned} f(p,q) &= p(1-p)[\alpha_1 - q(\alpha_1 + \alpha_2)] \\ g(p,q) &= q(1-q)[\beta_1 - p(\beta_1 + \beta_2)] \end{aligned}$$

where

$$\begin{aligned} \alpha_1 = a_{12} - a_{22}, \quad \alpha_2 = a_{21} - a_{11} \\ \beta_1 = b_{12} - b_{22}, \quad \beta_2 = b_{21} - b_{11}. \end{aligned}$$

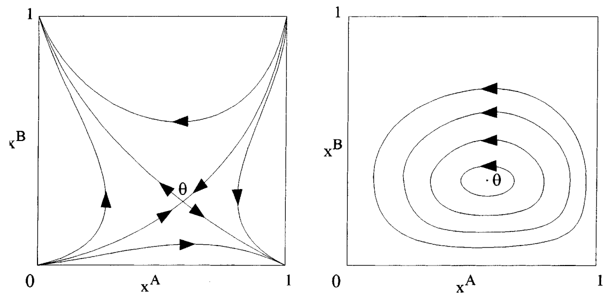


Figure 24: The solutions of the 2×2 replicator dynamics. The 'internally chain recurrent sets' for these differential equations are: the 5 singularities (for the flow on the left) and the entire phase space (for the flow on the right). The game corresponding to the replicator equation on the left is a coordination game and has three NE's (the points on the top left, bottom right, and the interior equilibrium point). The game associated to the picture on the right has only one NE, namely the interior equilibrium.

5.2.4 Comparison with replicator dynamics

Now compare (5.4) with the replicator dynamics equation

$$\begin{aligned} \dot{p}_i &= p_i[(Aq)_i - p \cdot Aq] \\ \dot{q}_j &= q_j[(Bp)_j - q \cdot Bp] \end{aligned}$$

where again the first convention is used. This also gives

$$\begin{aligned} \dot{p}_1 &= p_1[a_{11}q_1 + a_{12}q_2 - p_1(a_{11}q_1 + a_{12}q_2) \\ &\quad -p_2(a_{21}q_1 + a_{22}q_2)] \\ &= p_1(1-p_1)[\alpha_1 - q_1(\alpha_1 + \alpha_2)] \\ \dot{q}_1 &= q_1(1-q_1)[\beta_1 - p_1(\beta_1 + \beta_2)]. \end{aligned}$$

So

$$\dot{p}_1 = f(p_1, q_1), \dot{q}_1 = g(p_1, q_1)$$

is the two person replicator dynamics that we already encountered in Subsections 2.2 and 2.4.

As we saw there, the dynamics of this two person replicator system can be completely described. If there is an interior NE then there are two possibilities, see Figure 24, where the diagram on the right corresponds to a a game which is equivalent to a zero-sum game.

5.2.5 A formal connection with the replicator dynamics

Based on this, Posch (1997) and Hopkins & Posch (2005) showed the following:

Theorem 5.2. Consider the Arthur learning model in a two-player two strategy game. Then

- if the game has no strict Nash equilibrium and is equivalent to a zero sum game (as in Figure 24 on the right), then the learning algorithm has a continuum of asymptotically cycling paths. Almost all paths that are not asymptotically cycling converge either to the interior fixed point or to the boundary;
- if there is at least one strict Nash equilibrium and $C \geq a_{jk}, b_{jk}$ for $j, k = 1, 2$, then the learning algorithm a.s. converges to the set of strict Nash equilibria. All strict Nash equilibria are attained in the limit with positive probability.

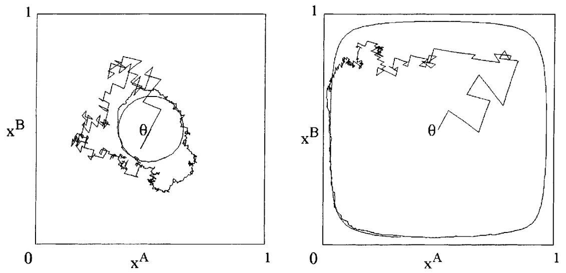


Figure 25: Two sample paths of the 2×2 reinforcement learning dynamics corresponding to the two systems considered in Figure 24.

In Figure 25, two runs of the learning process are drawn for a zero-sum game with a NE at $(1/2, 1/2)$ from which the runs where started. In fact, in this experiment the matrices $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ are chosen - these form a constant sum game, which has the same replicator orbits as the zero sum game with matrices $A - 1.5, B - 1.5$.

5.2.6 What happens if C is not large enough in Arthur's model?

Note that in the left panel of Figure 24, neither $(0,0)$ nor $(1,1)$ is a NE. Similarly, in the right panel of that figure none of the corner points are NE's. Nevertheless, if C is chosen insufficiently large, then the learning algorithm can reach these points with positive probability:

Proposition 5.1. Suppose that $0 < C < a_{k,l}, b_{k,l}$ for all k, l. Then

$$Prob\{\lim_{t\to\infty} p^t \to 1, \lim_{t\to\infty} q^t \to 1\} > 0.$$

Proof. Let us show that there is a positive probability that $p^t \to 1$. To do this, it is sufficient to show that $\prod_{t=1}^{\infty} p^t > 0$ because this implies that there is a positive probability that player A chooses action 1 forever.

Note that $\prod_{t=1}^{\infty} p^t > 0$ is equivalent to $\sum(1 - p^t) < \infty$. Since

$$\underline{p}^{t+1} = \underline{p}^t + \frac{u^t}{Ct + u^t}(\underline{x}^t - \underline{p}^t)$$

if player A chooses action 1 (which corresponds to $\underline{x}^t = e_1 = (1,0) = (1,0)$) at time t and player B action j then

$$p^{t+1} = p^t + \frac{a_{1j}}{Ct + a_{1j}}(1 - p^t).$$

So writing $d^t = 1 - p^t$ we get

$$d^{t+1} = d^t - \frac{a_{1j}}{Ct + a_{1j}} d^t.$$

Hence

$$\frac{d^{t+1}}{d^t} = 1 - \frac{a_{1j}}{Ct + a_{1j}}.$$

Since $a_{ij} > C$ for all i, j, there exists $\alpha, \alpha' > 1$ and t_0 so that for $t \ge t_0$

$$1 - \frac{\alpha}{t} < \frac{d^{t+1}}{d^t} < 1 - \frac{\alpha'}{t}.$$

This implies by the Raabe test that $\sum_{t=1}^{\infty} d^t$ converges. (The Raabe test states the following: assume $|c_n/c_{n+1}| \to 1$ and $n(|c_n/c_{n+1}|-1) \to R$. Then $\sum c_n$ converges if $R > 1$ and diverges if $R < 1$.)

Thus we have proved that if player A chooses action 1 all the time, then $\prod p^t > 0$. Thus it follows that there is indeed a positive probability that player A indeed chooses action 1 all the time. Since the same holds for player B the proposition follows. □

Exercise 5.3.
1. Show that the 'internally chain recurrent set' (as defined in the appendix) corresponding to the differential equations drawn in Figure 24 is as claimed in the caption of that figure.
2. In the previous proof $\frac{d^{t+1}}{d^t}$ does not really converge because the value of $\frac{d^{t+1}}{d^t} = 1 - \frac{a_{1j}}{Ct + a_{1j}}$ where a_{1j} depends on the action at time t. Show that nevertheless $\sum d^t$ converges. (Hint: it is enough to compare d^t with a sequence $\tilde{d}^t$ for which $\frac{\tilde{d}^{t+1}}{\tilde{d}^t} = 1 - \frac{\alpha'}{t}$ where $\alpha' > 1$.

5.3 The Erev-Roth model

In model (2) we have that $|\theta^t| \le |\theta^1| + tK$ where K is an upper bound for the utility of all actions. It follows that

$$p_i^t \ge \frac{\theta_i^1}{|\theta^1| + tK} \text{ and therefore } \prod_{t \ge 1}(1 - p_i^t) = 0.$$

(To see this implication, consider the logarithm of the product. We then need to prove $\sum_t \log(1 - p_i^t) \to -\infty$ and this follows from $\sum_t p_t^i \ge \sum_t \frac{\theta_i^1}{|\theta^1| + tK} = \infty$.) Note that $\prod_{t \ge 1}(1 - p_i^t) = 0$ implies that the probability of never choosing action i is zero,

5.3.1 The underlying differential equation

Following the same approach as in the Arthur model we now get, see Beggs (2005),

$$\begin{aligned} \dot{p}_i &= \frac{p_i}{a(t)}[(Aq)_i - p \cdot Aq] \\ \dot{a} &= -a + p \cdot Aq \\ \dot{q}_j &= \frac{q_j}{b(t)}[(Bp)_j - q \cdot Bp] \\ \dot{b} &= -b + q \cdot Bp \end{aligned} \tag{5.6}$$

Note that this is still quite close to the replicator system, but since a, b are not constant and are distinct there are subtle differences. It turns out that this implies that the solutions almost surely tend to Nash equilibria.

Theorem 5.3. Consider the Erev-Roth learning model in a two-player two strategy game. Then

- if the game has no strict Nash equilibrium and is equivalent to a zero sum game (as in Figure 24 on the right), then the learning algorithm converges to the interior fixed point;
- if there is at least one strict Nash equilibrium then the learning algorithm a.s. converges to the set of strict Nash equilibria. All strict Nash equilibria are attained in the limit with positive probability.

Exercise 5.4. 1. Determine the singularities of the adjusted equation (5.6).

2. Code up the previous algorithm in matlab or python and check whether the output of this code supports the previous theorem.

5.4 Q learning

This learning process was pioneered by Sutton & Barto (1998) and prior to that by Watkins & Dayan (1992). In this approach one additionally allows for the existence of distinct states. For example, these states could model which room in a house you are in. The actions are describes (for example) which door you exit a room. (Think of computer games.)

Suppose that you have a number of states $s \in S$ and a number of actions $a \in A$, and that you try to calculate the expected value of playing action i when in state s. Suppose that the vector $Q^t(s) \in \mathbb{R}^{\#A}$ where the a-th component $Q^t_a(s)$ of $Q^t(s)$ (where $a \in A$) is supposed to be an estimate of the future value being in state s and playing action a.

In Q learning one uses the following update rule. Suppose you are in state s, and moved to state s' and played action $a(t)$ then, taking $\alpha \in (0,1)$ and $\gamma \in (0,1)$ the update from time t to time $t+h$ is taken to be

$$Q^{t+1}(s) = Q^t(s) + \alpha \cdot \left(u^t + \gamma \max_{j \in A} Q^t_j(s') - Q^t(s) \cdot a(t) \right) a(t) \tag{5.7}$$

So only the component of Q^{t+1} which you play at time t is updated. This updating rule is called *Q-learning*. In our context $u^t_A = a^t \cdot Ab^t, \alpha, \vartheta \in (0,1)$.

Often α is called the *learning rate*, and γ the *discount factor*. The term $\max_j Q^t_j(s')$ should be understood as an estimate for the optimal future value of being in state s'. In these lecture notes we will assume that we are in a *one-state situation*, so that we can ignore s, and then (5.7) becomes

$$Q^{t+1} = Q^t + \alpha \cdot \left(u^t_A + \gamma \max_j Q^t_j - Q^t \cdot a(t) \right) a(t) \tag{5.8}$$

Note that the term in brackets is a scalar and the only the component i for which $a^t = e_i$ is updated in Q^t.

Remark 5.1. Q learning originates from the theory of Markov Decision Processes / Bandit problem in a random environment with stationary distribution (See Appendix)A6. In that case $Q^t(s)$ converges to a vector as $t \to \infty$ which is a fixed point of some Bellman equation. In that setting $\gamma =$ represents a discount factor on future earnings. In our setting γ makes little sense, so we will take $\gamma = 0$.

5.5 Various ways of choosing actions

Given vector Q (i.e. Θ^t or Q^t) how to choose action $a(t)$?

- **proportional to Q:** if all coordinates of Q are positive, one can choose as before

$$p(Q) = \frac{Q}{|Q|_1}.$$

- **ϵ-greedy choice:** according to the probability vector

$$p(Q) = (1-\epsilon)\mathcal{B}R_I(Q) + \epsilon(1/n, \ldots, 1/n)$$

  Note that $\mathcal{B}R_I(Q)$ is the unit vector corresponding to the largest component of Q (plus tie rule).

- **softmax:** according to the probability vector

$$\text{softmax}_\tau(Q) = \frac{1}{\sum_i \exp(\tau Q_i)}(\exp(\tau Q_1), \ldots, \exp(\tau Q_n)).$$

  - $\tau \downarrow 0$: uniform distribution $(1/n, \ldots, 1/n)$ and
  - $\tau = \infty$ then this puts full weight on the largest component of Q.

  A small value for $\tau > 0$ corresponds to 'exploration' and a large value for τ correponds to 'exploitation'.

5.6 Q-Learning with softmax

In this subsection we will consider a related model, called *frequency adjusted Q-learning*:

$$Q^{t+1} = Q^t + \alpha\left(u^t + \gamma \max_j Q_j^t \cdot \mathbb{1} - Q^t\right) \tag{5.9}$$

while choosing actions according to the softmax vector $x(t) = (x_1(t), \ldots, x_n(t))$

$$x_i(t) = \frac{e^{\tau Q_i^t}}{\sum_j e^{\tau Q_j^t}}, i = 1, \ldots, n. \tag{5.10}$$

Here $\mathbb{1}$ is the vector with all components equal to 1 and u^t is the conditional expected payoff *vector* where u_i^t is the payoff you would receive if you chose action i given all the information that is currently available about the other player. Note that in (5.9) you do not only update the i-th component of Q^r (where $e_i = a^t$) but also hypothetically consider *all* actions and adjust Q at each time according to the payoff you expect from these actions.

Instead of taking the time step $h = 1$ as in the updating rule (5.9), we will here show that consider small time steps $h > 0$. Indeed, we will show that $h \to 0$ you obtain a differential equation which is closely related to the replicator system. So consider the analogue

$$Q^{t+h} = Q^t + \alpha h\left(u^t + \gamma \max_j Q_j^t \cdot \mathbb{1} - Q^t\right) \tag{5.11}$$

of (5.9). Using (5.10) we get

$$\frac{x_i(t+h)}{x_i(t)} = \frac{e^{\tau Q_i^{t+h}} \sum_j e^{\tau Q_k^t}}{e^{\tau Q_i^t} \sum_j e^{\tau Q_j^{t+h}}} = \frac{e^{\tau \Delta Q_i^t}}{\sum_j x_j(t) e^{\tau \Delta Q_j^t}}$$

where $\Delta Q_i^t = Q_i^{t+h} - Q_i^t$ and where we used $e^{\tau Q_j^t} e^{\tau \Delta Q_j^t} = e^{\tau Q_j^{t+h}}$. This gives

$$x_i(t+h) - x_i(t) = x_i(t) \left(\frac{e^{\tau \Delta Q_i^t} - \sum_j x_j(t) e^{\tau \Delta Q_j^t}}{\sum_j x_j(t) e^{\tau \Delta Q_j^t}} \right).$$

Let us now consider the limit $h \to 0$. Since $\sum_j x_j(t) = 1$ as x is a probability vector and $e^{\tau \Delta Q_k(t)} \to 1$ as $h \to 0$, the denominator of the previous expression tends to 1 as $h \to 0$. So

$$\lim_{h \to 0} \frac{x_i(t+h) - x_i(t)}{h} = x_i(t) \lim_{h \to 0} \left(\frac{e^{\tau \Delta Q_i^t} - \sum_j x_j(t) e^{\tau \Delta Q_j^t}}{h} \right).$$

Using $e^x = 1 + x + O(x^2)$ and $\sum x_j = 1$ this gives

$$\frac{dx_i}{dt} = x_i \tau \left(\frac{dQ_i}{dt} - \sum_j \frac{dQ_j}{dt} x_j \right). \tag{5.12}$$

From (5.11) we obtain

$$\frac{dQ_i}{dt} = \alpha \left(u_i^t + \gamma \max_j Q_j - Q_i \right).$$

Since $\sum x_j = 1$, substituting this in the equation (5.12) for $\frac{dx_i}{dt}$ the term with γ drops out and we obtain

$$\begin{aligned} \frac{dx_i}{dt} &= x_i \tau \alpha \left(u_i^t - Q_i - \sum x_j u_j^t + \sum_j Q_j x_j \right) = \\ &= x_i \tau \alpha \left(u_i^t - \sum x_j u_j^t + \sum_j ((Q_j - Q_i) x_j) \right) \end{aligned}$$

Notice that $\frac{x_j}{x_i} = \frac{e^{\tau Q_j}}{e^{\tau Q_i}}$ and so $\sum_j x_j \log(x_j/x_i) = \tau \sum_j x_j (Q_j - Q_i)$. Thus we get

$$\frac{dx_i}{dt} = x_i \tau \alpha \left(u_i^t - \sum x_j u_j^t + (1/\tau) \sum_j x_j \log(x_j/x_i) \right).$$

If the pay-off matrices of player I and II are A and B then, since u_i^t is the payoff that you would obtain from choosing action i and the other player is expected to play actions according to her vector y, the above expression gives

$$\begin{aligned} \frac{dx_i}{dt} &= x_i \tau \alpha \left((Ay)_i - x \cdot Ay + (1/\tau) \textstyle\sum_j x_j \log(x_j/x_i) \right) \\ \frac{dy_i}{dt} &= y_i \tau \alpha \left((Bx)_i - y \cdot Bx + (1/\tau) \textstyle\sum_j y_j \log(y_j/y_i) \right) \end{aligned} \tag{5.13}$$

Note that $\sum_i \frac{dx_i}{dt} = 0$ because $\sum_i x_i(Ay)_i = x \cdot Ay$ and because

$$\sum_i \sum_j x_i x_j (\log(x_j/x_i)) = 0$$

since $x_i x_j \log(x_j/x_i) + x_j x_i \log(x_i/x_j) = 0$. Of course (5.13) can also be written as

$$\begin{aligned} \frac{dx_i}{dt} &= x_i \tau\alpha \left((Ay)_i - x \cdot Ay + (1/\tau)\left[-\log x_i + \textstyle\sum_j x_j \log x_j\right]\right) \\ \frac{dy_i}{dt} &= y_i \tau\alpha \left((Bx)_i - y \cdot Bx + (1/\tau)\left[-\log y_i + \textstyle\sum_j y_j \log y_j\right]\right) \end{aligned} \tag{5.14}$$

Remember that τ is the coefficient in

$$x_i(t) = \frac{e^{\tau Q_i^t}}{\sum_j e^{\tau Q_j^t}}, i = 1, \ldots, n.$$

Remember that as $\tau \to \infty$ then $x(t) \to e_m$ where m is the largest component of Q. Notice that when $\alpha = 1/\tau \to 0$ (i.e. when $\tau \to \infty$ and one considers the ‘exploitation’ limit) then (5.14) converges to the usual replicator dynamics. This is why the term in square brackets in (5.13) and (5.14) is called the ‘explorationrq term.

Exercise 5.5. 1. Consider the following games:

$$\text{prisoner dilemma} : \begin{pmatrix} (1,1) & (5,0) \\ (0,5) & (3,3) \end{pmatrix} \qquad \text{battle of the sexes} : \begin{pmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{pmatrix}$$

$$\text{matching pennies:} \begin{pmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{pmatrix}$$

In this section we have seen three ways in which players could learn to play these games, namely (5.8), (5.9) and (5.14) where in the first two the vectors $x(t), y(t)$ are determined by the vectors $Q^A(t), Q^B(t)$ through equation (5.10). The solutions are two of these learning algorithms are drawn on the next page. Can you replicate these figures through a simulation in matlab or python? (The figures labeled "FAQ" are associated to yet another learning algorithm.)

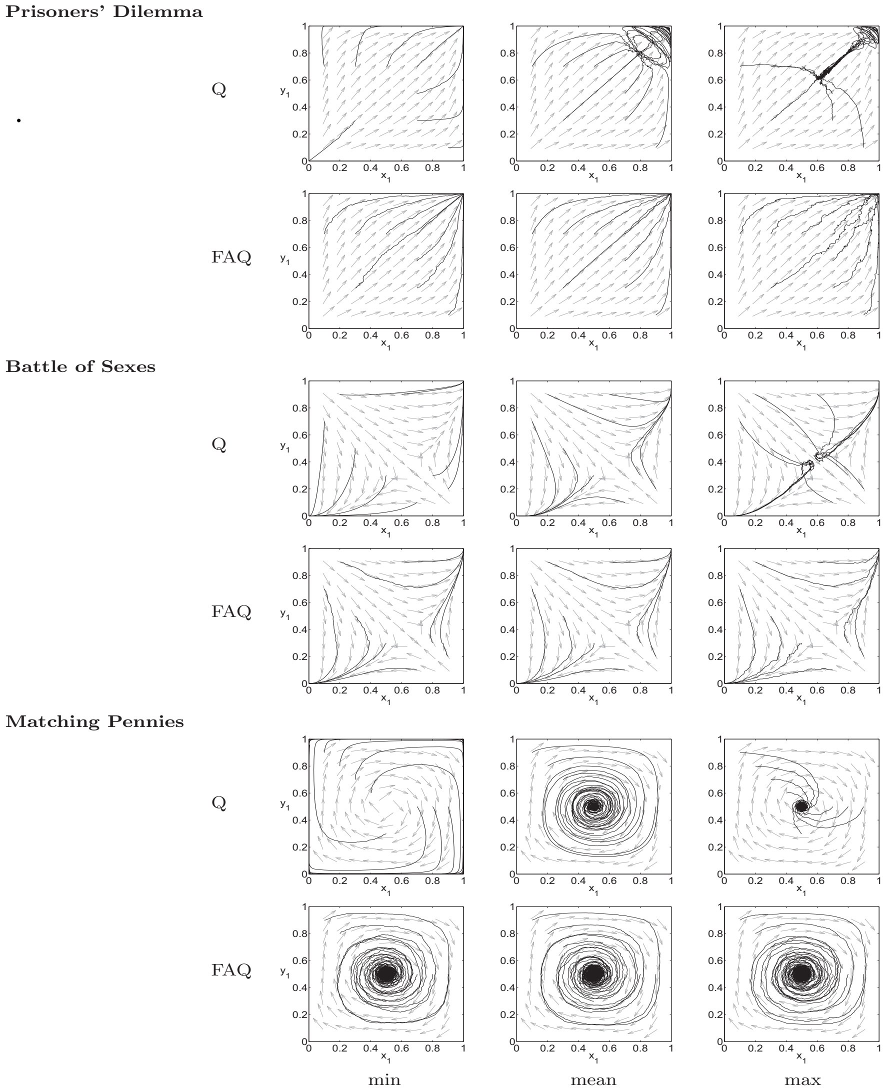


Figure 4: Comparison of Q-learning to FAQ-learning with various Q-value initializations in the Prisoners' Dilemma, the Battle of Sexes and Matching Pennies. The Q-values are initialized centered at the minimum (left), mean (center) and maximum (right) possible Q-value given the reward space of the game.

314

5.7 So what is the message?

Learning theory is a very hot topic right now. Much of the more practical work is about finding suitable coefficients which 'work'. This is not so surprising in view of the theoretical results given before.

The dynamics of the learning models we have considered in this section are all somewhat related to the replicator dynamics. We also have seen that replicator dynamics in 3×3 games can have chaotic dynamics. This suggests that one should not expect convergence within learning algorithms in a setting where two players compete. This is the topic for ongoing research.

5.8 Some computer experiments: what if the opponent has a different strategy?

Consider the following game: $\begin{pmatrix} 2,1 & 1,2 \\ 1,2 & 2,1 \end{pmatrix}$.

Suppose the 2nd player uses

1. fictitious play;
2. takes a (myopic) best response to player 1's current action;
3. plays the minmax strategy.

Then player 1's average payoff converges rapidly to 1.5. Indeed, Beggs [2005] did some computer simulations. Against each opponent the ER rule was run 100 times in a run of length 10,000, with initial reinforcements (1,1.5).

The mean average payoff was

1. 1.48 st.dev. 0.04
2. 1.49 st.dev. 0.01
3. 1.5 st.dev. 0.003

6 The best response dynamics

In addition to the replicator dynamics (which was proposed in the 1980's), another game dynamics is often studied. In this dynamics, a player's mixed strategy evolves at each moment towards the best strategy: define as before

$$\mathcal{B}R(x) = \arg\max_y y \cdot Ax$$

and let

$$\dot{x} = \mathcal{B}R(x) - x. \tag{6.1}$$

This differential equation is called the best response dynamics and was proposed in the 1950's and so predates replicator dynamics by several decades. In the next chapter we will explain how the best response dynamics is closely related to a very natural learning dynamics.

Note that $\mathcal{B}R(x)$ is a non-empty convex set, and so strictly speaking (6.1) is not a differential equation but rather a differential inclusion

$$\dot{x} \in \mathcal{B}R(x) - x. \tag{6.2}$$

and so we cannot apply the usual existence and uniqueness results to this equation. Fortunately $x \mapsto \mathcal{B}R(x)$ is upper semi-continuous. (This means that for each closed set K the set $\{x; \mathcal{B}R(x) \cap K \neq \emptyset\}$ is closed.) It turns out that this implies for each $x_0 \in \Delta$ there exists

- a continuous curve $t \mapsto x(t)$ with $x(0) = x_0$ so that
- $t \mapsto x(t)$ almost everywhere differentiable and so that
- $\dot{x}(t) = \mathcal{B}R(x(t)) - x(t)$ for each t at which $t \mapsto x(t)$ is differentiable.

When these properties are satisfied we call $t \mapsto x(t)$ a solution of (6.1).

In this chapter we will only encounter the following situation: for each (continuous) solution $\mathbb{R} \ni t \mapsto x(t)$ of (6.1) there exists a countable set $I \subset \mathbb{R}$ without accumulation points so that

$$\mathcal{B}R(x(t)) \text{ is a is single-value for each } t \in \mathbb{R} \setminus I.$$

Write $\mathbb{R} \setminus I = \cup_j (t_j, t_{j+1})$ with $t_j < t_{j+1}$ for all $j \in \mathbb{Z}$. By definition for $t \in (t_j, t_{j+1})$ we have that $\mathcal{B}R(x(t))$ is single-value which means that $Ax(t)$ has a single largest component, and so there exists $i(j) \in \{1, \dots, n\}$ so that $\mathcal{B}R(x(t)) = e_{i(j)}$. Hence (6.1) takes the form

$$\dot{x} = e_{i(j)} - x \text{ for } t \in (t_j, t_{j+1}).$$

This means that $x(t)$ moves along the straight line through $e_{i(j)}$ while $t \in (t_j, t_{j+1})$. For $t \in \{t_j, t_{j+1}\}$ we have that $\mathcal{B}R(x(t))$ is multi-valued, and so $x(t)$ lies in an indifference line. Then for $t \in (t_{j+1}, t_{j+2})$ the orbit again moves along a straight line, but now pointing towards $e_{i(j+1)}$.

Let us consider a matrices for which we already studied the replicator dynamics.

6.1 Rock-scissor-paper game and some other examples

Example 6.1 (Rock-scissor-paper)**.** Consider the matrix

$$A = \begin{pmatrix} 0 & -b & a \\ a & 0 & -b \\ -b & a & 0 \end{pmatrix}, \quad Ax = \begin{pmatrix} ax_3 - bx_2 \\ ax_1 - bx_3 \\ ax_2 - bx_1 \end{pmatrix} \tag{6.3}$$

with $a, b > 0$ similar to the one for which we already consider the replicator ode in Example (1.11). Remember that this system has a unique Nash equilibrium at $E = (1/3)\mathbb{1}$. Note that $\mathcal{B}R(e_i) = e_{i+1}$ and that $\mathcal{B}R(E) = \Delta$ and that $\mathcal{B}R(x)$ takes values e_1, e_2, e_3 outside the indifference line segments $(Ax)_i = (Ax)_j$ and along those line segments is multivalued. For example $(Ax)_2 = (Ax)_3$ corresponds to $ax_1 - bx_3 = ax_2 - bx_1$ and so when $x_3 = 0$ this means $x_1 = a/(a+b)x_2$.

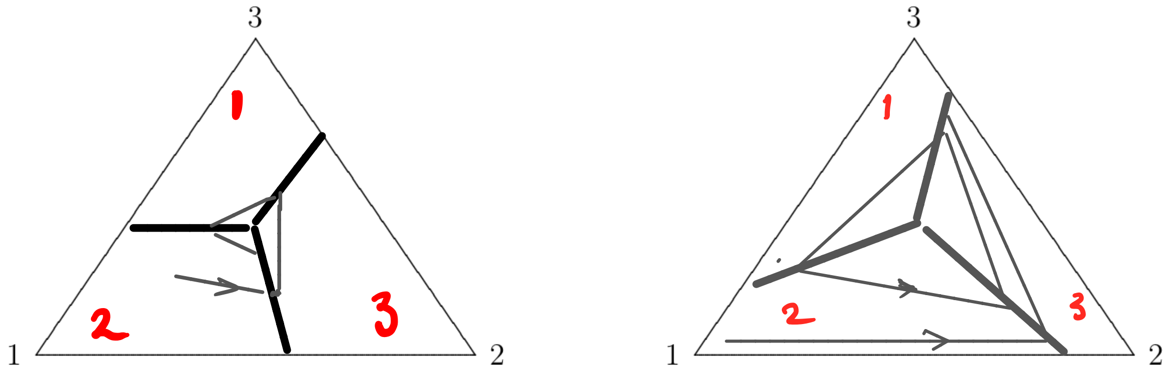


Figure 26: The best response dynamics corresponding to (6.3). Solutions consist of piecewise straight lines, directed at e_2, e_3, e_1 etc. On the left the case when $a > b$ is drawn and on the right the case when $a < b$ and when solutions cycle to the Shapley triangle.

Take $V(x) = \max_i (Ax)_i$. Then $V(x) = e_i \cdot Ax$ where $e_i = \mathcal{B}R(x)$ is piecewise constant.

- Note that V is continuous (and differentiable outside the lines where $\mathcal{B}R(x)$ is multi-valued).
- Note that $V(x) \geq V(E)$ for all $x \in \Delta$. Indeed, write $x = z + E$ with $\sum z_i = 0$. Then $V(x) = V(E) + V(z)$. Moreover, the definition of A implies $\sum_i (Az)_i = 0$ and so $V(z) = \max(Az)_i \geq 0$ except if $(Az)_1 = (Az)_2 = (Az)_3 = 0$ which only holds if $z = 0$.

Also note that since $A_{ii} = 0$ for $i = 1, 2, 3$, in the interior of the region where $BR(x) = e_i$ we have

$$\dot{V} = e_i \cdot A \cdot \dot{x} = e_i \cdot A(e_i - x) = -e_i \cdot Ax = -V$$

except at the Nash equilibrium E. It follows that $V(x(t)) = V(x(0))e^{-t}$ and at first sight this seems to suggest that $V(x(t)) \to 0$ as $t \to \infty$. However, since $V(x) \geq V(E)$ for all $x \in \Delta$ this may not be possible. Indeed, $V(E) = (a-b)/3$.

When $a > b$ then $V(E) > 0$. It follows that orbits reach E in *finite time*.

When $a < b$ then $V(E) < 0$ and so solution starting outside E do NOT converge to E but to the set where $V = 0$, which is a triangle, called the Shapley triangle.

Example 6.2. Let us consider the $\mathcal{B}R$-dynamics associated to the matrix $A = \begin{pmatrix} 0 & 6 & -4 \\ -3 & 0 & 5 \\ -1 & 3 & 0 \end{pmatrix}$ we considered before in Example 1.8. Again the NE is $E = (1/3)\mathbb{1}$. The best response regions are again determined by drawing the lines Z_{ij} lines. As before, the $\mathcal{B}R$-dynamics is multivalued along these indifference lines but there is a difference compared to the previous example.

For example, along the segment of the line $Z_{1,3}$ where regions 1 and 3 meet the 'flow' is non-continuous, see the figure below: if you just below and near this line then you flow further down, and if you are just above then you flow further up.

Along the segment of the line $Z_{2,3}$ where the regions 2 and 3 meet, the opposite is true. If you are just above this segment then you flow down, and if you just below then you flow up. So the flow near this segment 'pushes' you towards the segment. One can formalise this argument to show that one a uniquely defined 'semi-flow' near this segment: if you are on it, then you flow towards E and if you are near this segment then you flow towards it in finite time, and once you hit it then you flow towards E, again hitting it in finite time.

We will not try to formalise this argument properly in these lectures.

Exercise 6.1. 1. Consider the matrix A from Example 6.2. Take as before $V(x) = \max_i Ax$. Show that since $A_{ii} = 0$ for $i = 1, 2, 3$ we still have $\dot{V} = -V$ on the interior of the regions where $\mathcal{B}R$ is constant. One can show that the level set of $V = 0$ in Δ is no longer is a triangle, but is as shown in the blue set drawn in the top right in Figure 27. (You are not asked to calculate the position of these lines in detail.) Show that

(a) the blue set consists of pieces of three pieces of straight lines going through e_1, e_2, e_3.

(b) the segments of the blue set that are contained in the interior of the regions where $\mathcal{B}R$ is constant, have the property that if you start on these segments then you stay on this segment (until you hit an indifference line).

(c) that since $A_{ii} = 0$ for $i = 1, 2, 3$ we still have $\dot{V} = -V$ on the interior of the regions where $\mathcal{B}R$ is constant.

(d) This suggests that V decays exponentially fast to zero. Show that this is misleading, because $\dot{V} = -V$ no longer holds on $Z_{2,3}$.

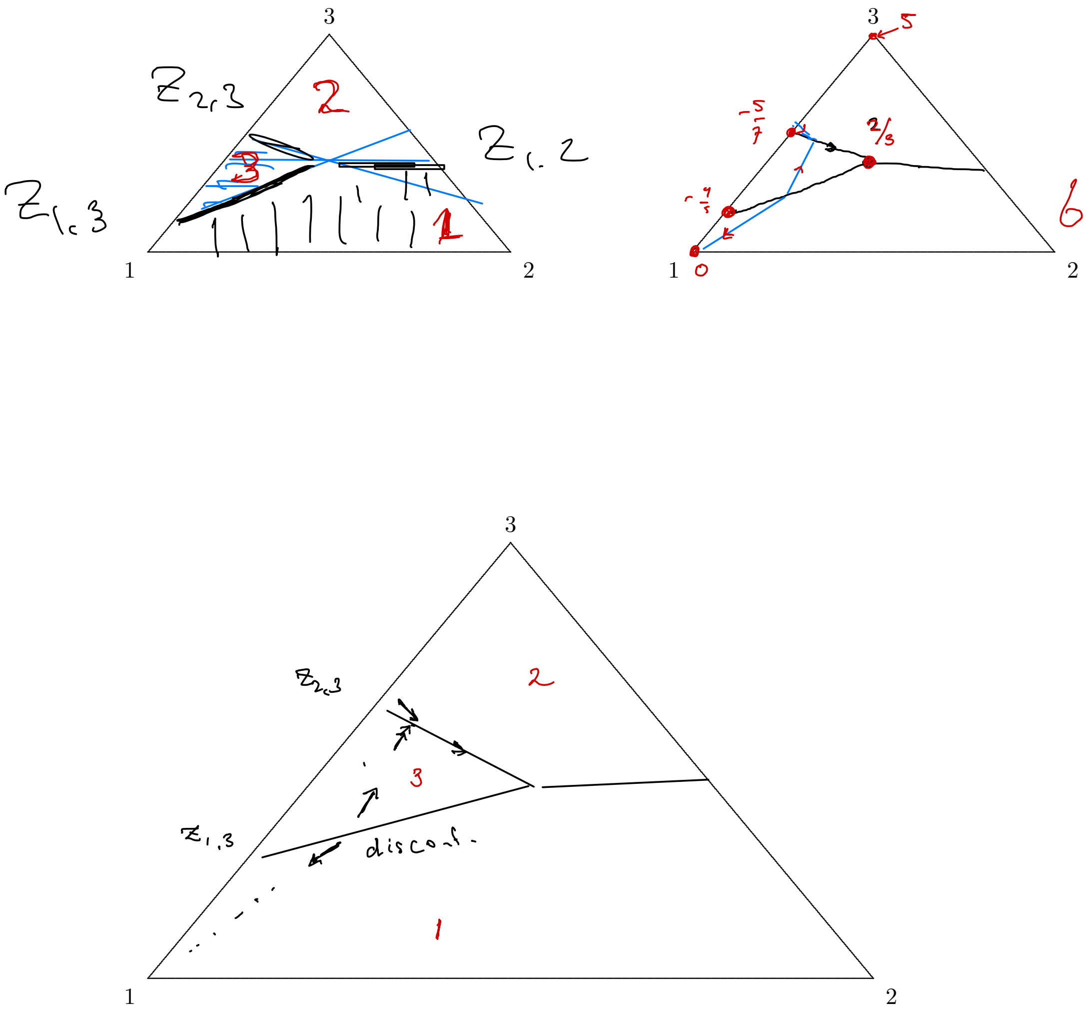


Figure 27: The best response dynamics corresponding to Example (6.2). Note that the solutions do always not depend continuously on the initial conditions. For example, along the line Z_{13} solutions just south of this line head towards e_1 and just to the north to e_3. Along the line itself, the solution is not uniquely defined. On the other hand, near $Z_{2,3}$ one can 'extend' so that it becomes a continuous 'semi-flow'.

6.2 Two player best response dynamics

The best response dynamics corresponding to two populations is

$$\begin{aligned}\dot{x} &= \mathcal{B}R_A(y) - x \\ \dot{y} &= \mathcal{B}R_B(x) - y\end{aligned}$$

Example 6.3. Let us consider the example of $\begin{pmatrix} (-1,1) & (0,0) \\ (0,0) & (-1,1) \end{pmatrix}$, where obviously we use the 2nd convention from equation (2.2. Here both players have opposite interests (the sum of the payoff's is always zero) and there is a unique interior NE, namely at $E = (1/2, 1/2) \times (1/2, 1/2)$. Let us show that in this case solutions go to this NE. Take

$$V(x, y) = \mathcal{B}R_A(y)\, Ay + x \cdot B\ \mathcal{B}R_B(x).$$

This function V is continuous because $\mathcal{B}R_A(y)\, Ay = \max_i (Ay)_i$ and maximum of several continuous functions is again a continuous function and similarly for $x{\cdot}B\,\mathcal{B}R_B(x) = \max_j (x^{tr}B)_j$. Notice

$$\mathcal{B}R_A(y)Ay \geq xAy \text{ and } x \cdot B\mathcal{B}R_B(x) \geq xBy = -xAy.$$

It follows that $V(x, y) \geq 0$. Moreover, at $E = (E^A, E^B)$ we have $V(E) = \mathcal{B}R_A(E^B) \cdot AE^B + E^A \cdot B\mathcal{B}R_B(E^A) = E^A \cdot AE^B + E^A \cdot BE^B = 0$. Moreover,

$$\begin{aligned}\dot{V} &= \mathcal{B}R_A(y) \cdot A\dot{x} + \dot{x} \cdot B\,\mathcal{B}R_B(x) \\ &= \mathcal{B}R_A(y) \cdot A(\mathcal{B}R_B(x) - y) + (\mathcal{B}R_A(y) - x) \cdot B\,\mathcal{B}R_B(x) \\ &= -V\end{aligned}$$

where in the last step we used $A + B = 0$. It follows that $V(x(t), y(t)) = e^{-t}V(x(0), y(0))$. This means that orbits tend exponentially fast to the Nash equilibrium E. The best response orbits spiral to the NE.

Example 6.4. The zero-sum bimatrix game $\begin{pmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{pmatrix}$, corresponds to the matching pennies game. The same analysis as above show that the NE is again $E = (1/2, 1/2) \times (1/2, 1/2)$ and that the BR solutions spiral to this NE, but in the opposite direction as in the game from Example 6.3. (Draw the phase diagram.)

Example 6.5. The zero-sum bimatrix game $\begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (-1,1) \end{pmatrix}$ has $NE = (e_1, e_2)$. Show that all the BR solutions tend to this NE which lies on the corner of the state space. (Draw the phase diagram.)

Exercise 6.2. Do solutions in Example 6.3 take an infinite amount of time to reach E? Why is it the case that in Example 4.1 solutions reach E in finite time? Note that in both cases, the speed does not go to zero (as would be the case for a singularity of a smooth ODE).

6.3 Convergence and non-convergence to Nash equilibrium for Best Response Dynamics

One of the main reasons best response dynamics was introduced in the 50's is that it was expected that it would provide a way to find a Nash Equilibrium. In other words, that the dynamics would always converge to the Nash equilibrium, and thus this dynamics would provide a mechanism for players to evolve towards a Nash equilibria (or to the set of Nash equilibria). For zero-sum games this is indeed the case. Indeed, the argument given in the previous example generalises to:

Theorem 6.1. Assume that (A, B) is a zero-sum game. Then the best response dynamics and also the FP dynamics (introduced in the next section) converges to the set of Nash equilibria of the game.

In fact, one has convergence to Nash equilibria for 2×2 and $2 \times n$ games and several other classes of games.

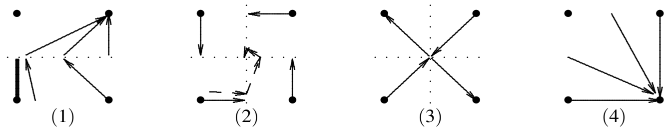


Figure 28: The possible motions in 2×2 games (up to relabeling, and shifting the indifference lines (drawn in dotted lines).

Example 6.6 (Two by two games)**.** For general 2×2 game there are only 4 types of best response dynamics (up to re-labelling the axis), see Figure 28. For example, when $\begin{pmatrix} (1,1) & (0,0) \\ (0,0) & (1,1) \end{pmatrix}$ there are three Nash equilibrium, namely $E = ((1/2, 1/2), (1/2, 1/2))$ and $((0,0),(0,0))$ and $((1,1),(1,1))$. The orbits are then as in Figure 28 subfigure 3. Can you find matrices A, B so that the dynamics is as in n Figure 28 subfigure 1?

However, in general one does not have convergence. For example:

Example 6.7 (Shapley system)**.** Take

$$A_\beta = \begin{pmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{pmatrix} \quad B_\beta = \begin{pmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{pmatrix}, \tag{6.4}$$

where we use the 2-nd convention.

Note that (E^A, E^B) where $E^A := (1/3, 1/3, 1/3)$ and $E^B := (1/3, 1/3, 1/3)'$ is the Nash equilibrium. (How can one work out that there are no other Nash equilibria?).

For $\beta = 0$ this corresponds to the situation that $A = Id$ so player one wants to copy what player two is doing ($BR_A(e_i) = e_i$), and B prefers 3, 2, 1 when player A plays 1, 3 and 2, so player B want so do something different from player A (because $BR_B(e_i) = e_{i-1}$). This games was introduced by the Nobel prize winner Shapley in 1964, to show that the dynamics of FP does not necessarily converge to a Nash equilibrium, but to a periodic orbit.

Lemma 6.1 (Shapley)**.** For $\beta = 0$ there exists a periodic orbit $\gamma : \mathbb{R} \to \Delta \times \Delta \subset \mathbb{R}^6$.

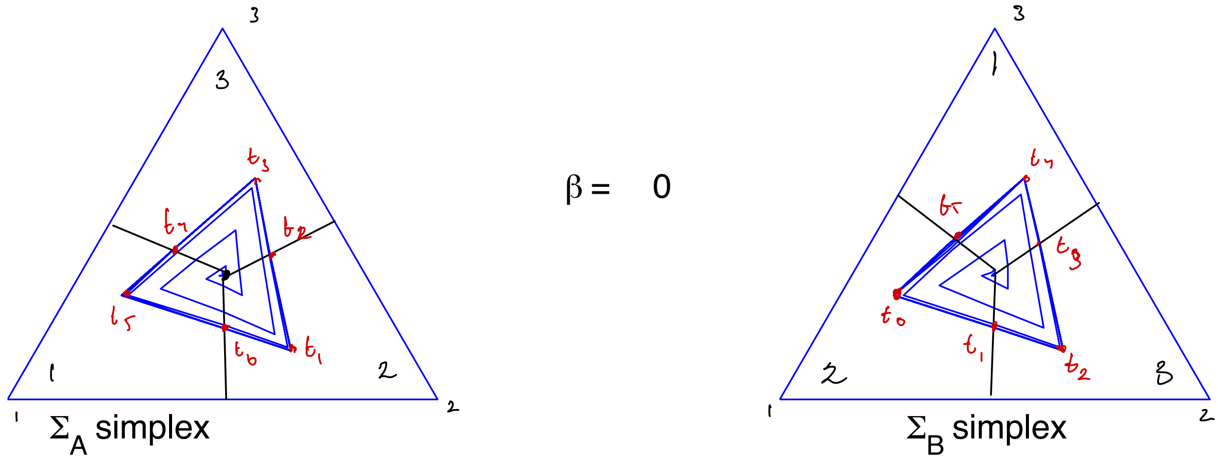


Figure 29: Shapley's periodic orbit for the best response dynamics of (6.4) for $\beta = 0$

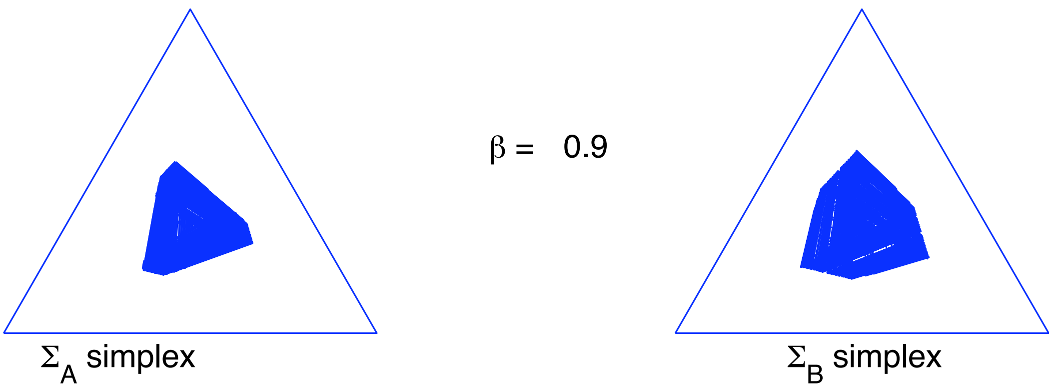


Figure 30: The motion for the best response dynamics of the Shapley system (6.4) for $\beta = 0.9$. Note that these are projections of the orbit in the four-dimensional space onto the two simplices. The dynamics appears to be chaotic. It was rigorously proved that there are infinitely many periodic orbits and 'horseshoes' in this dynamical system.

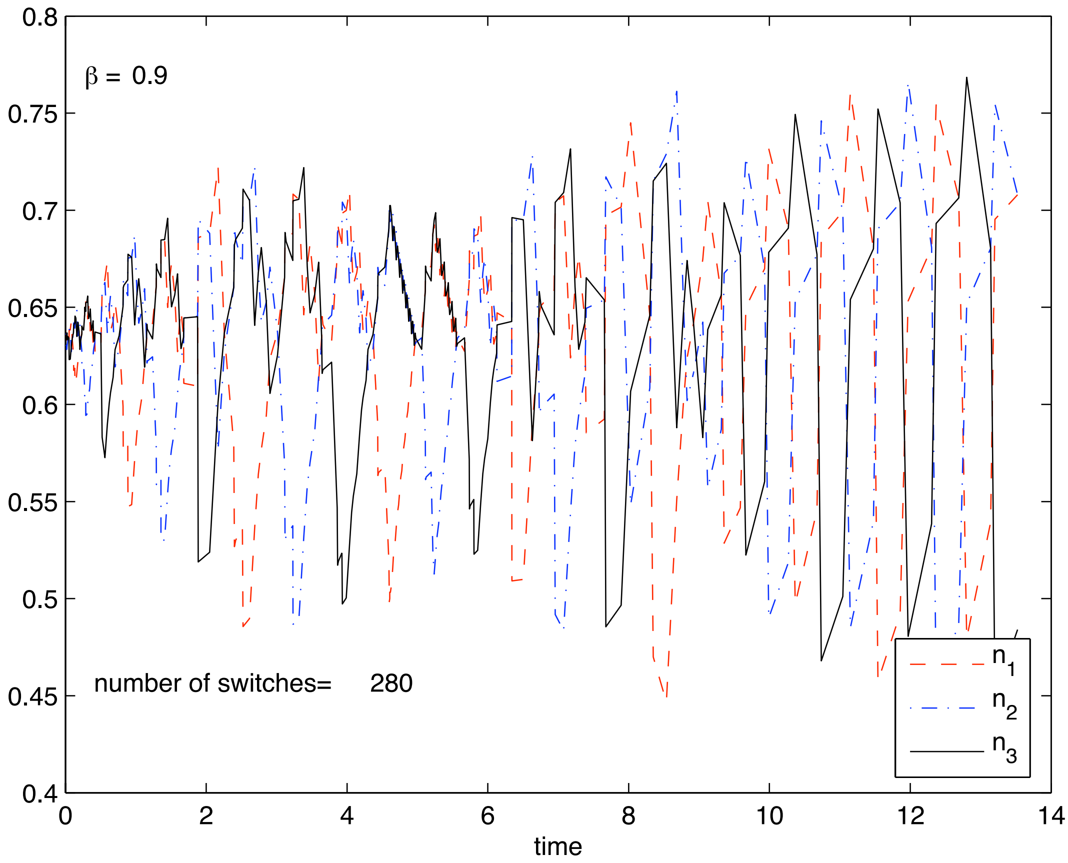


Figure 31: The three components of payoff vector $n(t) = Ay(t)$ as a function of time when we take $\beta = 0.9$ in the Shapley system (6.4).

Proof. Let us first explain what this periodic orbit γ will look like before proving its existence. We have $BR_A(e_i) = e_i$ and $BR_B(e_i) = e_{i+1}$ (note that we using the 2nd notation for the matrices). Let π_A, π_B be the projections of $\Delta_A \times \Delta_B \subset \mathbb{R}^6$ onto the two triangles shown in Figure 29. The triangles drawn in this figure correspond to the projections $\pi_A(\gamma)$ and $\pi_B(\gamma)$ of γ. Let T be the period of γ and let $0 = t_0 < t_1 < \cdots < t_5 < t_6 = T$ be the times when $\pi_A(\gamma(t))$ or $\pi_B(\gamma(t))$ are contained in one of the indifference lines. Note that when $\pi_A(\gamma(t))$ lies in an indifference line at $t = t'$, then $t \mapsto \pi_B(\gamma(t))$ changes from moving towards one corner for $t < t'$ close to t' to moving towards another corner for $t > t'$ close to t'. In fact, for each t we have that $\gamma(t)$ intersects at most one of the indifference lines. The points $\pi_A(\gamma(t_i))$ and $\pi_B(\gamma(t_i))$ are indicated in the figure, and note that the points move anti-clockwise in the triangles and head towards e_i in Δ_A when $\gamma(t)$ is in the region in Δ_B marked with i (and vice versa). The curve γ is a solution of the piecewise smooth BR dynamics, but for $t \in (t_i, t_{i+1})$ this simply reduces to

$$\dot{\gamma}(t) = (e_i, e_j) - \gamma(t) \text{ for } t \in (t_i, t_{i+1}). \tag{6.5}$$

Here (e_i, e_j) are best response choices. i.e., $e_i = BR_A(\pi_A(\gamma)$ and $e_j = BR_B(\pi_B(\gamma(t))$. The solution of (6.5) is

$$\gamma(t) = (1 - e^{-t})(e_i, e_j) + e^{-t}\gamma(0) \text{ for } t \in (t_i, t_{i+1}). \tag{6.6}$$

So $(t_i, t_{i+1}) \ni t \mapsto \gamma(t)$ is a straight line in $\mathbb{R}^6$ (which is contained in $\Delta_A \times \Delta_B$). Moreover, for i's you obtain a different straight line. Therefore $\gamma(t)$ will be a closed continuous curve consisting of 6 straight lines.

To see that there exists indeed a periodic orbit γ requires an explicit calculation. In Shapley's original paper *Some topics in two-person games, 1963*, he simply states that the corners of the hexagon corresponding to the periodic orbit γ are

$$\gamma(t_0) = (\theta^3, \theta^3, \theta, \theta^4, 1, \theta^2)/C$$

$$\gamma(t_1) = (\theta^2, \theta^4, 1, \theta^3, \theta^3, \theta)/C$$

$$\gamma(t_2) = (\theta, \theta^3, \theta^3, \theta^2, \theta^4, 1)/C$$

(and continuing this cyclically for the other points $\gamma(t_i)$) where θ is the unique real root $\theta^3 - \theta^2 = 1$ (which is $\theta > 1$ and in fact $\theta \approx 1.466$) and where C is chosen so that these points are in $\Delta_A \times \Delta_B$. Indeed, $C = 2\theta^3 + \theta = 1 + \theta^2 + \theta^4$; the last equality holds because of $\theta^3 - \theta^2 = 1$. That this polygon indeed corresponds to the solution of the $\mathcal{B}R$ dynamics can be shown by an explicit calculation using equation (6.6). Indeed, if we take $\lambda = (\theta - 1)/\theta \in (0, 1)$ then using the definition of C,

$$\lambda C = (\theta - 1)(2\theta^2 + 1) = 2\theta^3 - 2\theta^2 + \theta - 1$$

and

$$(1 - \lambda)\theta = 1, (1 - \lambda)\theta^3 + \lambda C = \theta^4, (1 - \lambda) + \lambda C = \theta^3$$

and from this we obtain that the line from $\gamma(t_0)$ to (e_2, e_2) passes through $\gamma(t_1)$:

$$\gamma(t_0) + \lambda((e_2, e_2) - \gamma(t_0)) = \gamma(t_1).$$

and similarly

$$\gamma(t_1) + \lambda((e_3, e_2) - \gamma(t_1)) = \gamma(t_2).$$

So if take

$$1 - e^{-t_1} = \lambda \text{ and } e^{-t_1} = 1 - \lambda$$

and

$$1 - e^{-(t_2-t_1)} = \lambda \text{ ands } e^{-(t_2-t_1)} = 1 - \lambda$$

we see that indeed these points are orbits of a solution of (6.5) and (6.6). Using the symmetry of the equation we obtain the full periodic orbit. In Appendix A, Harris & Sparrow & SvS 2008 the existence of a periodic orbit for $\beta \in (-1, 1)$ is shown through a calculation. An abstract argument which does not require calculations is given for $\beta \in (-1, 0]$ in SvS & Sparrow 2011, Proposition 3.1.

□

For $\beta = \phi$ where ϕ is the *golden number* (i.e. $\phi := (\sqrt{5} - 1)/2 \approx 0.618$), the game is equivalent to a zero-sume game (rescaling B to $\tilde{B} = \phi(B - 1)$ gives $A + \tilde{B} = 0$). Hence in this case by Theorem 6.1 play always converges to the interior equilibrium (E^A, E^B).

For $\beta \in (\phi, \tau)$ where $\tau \approx 0.915$ the dynamics is chaotic, as is shown in Sparrow & SvS 2011.

Exercise 6.3.

1. Show that the best response associated to a 3×3 matrix A is the same as that associated to $A' = cA + \begin{pmatrix} \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \end{pmatrix}$ provided $c > 0$.

2. Show that when β is equal to the golden mean, the game defined in (6.4) is indeed (equivalent) to a zero sum game. (We say that the games are equivalent if the best response dynamics is the same.)

3. Go through the argument in the previous lemma in detail and show that we indeed obtain a periodic orbit.

7 Fictitious play: a learning model

There are several models for learning which aim to model human behaviour while others are aimed at providing efficient algorithms for computing various generalisations of the Nash equilibrium. Some models have their roots in economics, whereas others in computer science literature. In the remainder of this course we will discuss some of the main models:

- fictitious play (many people, starting with Brown and Robinson in the 50's), Fudenberg, Levine,....
- reinforcement learning Bush and Mosteller (1951, 1955), (Roth, Erev, Arthur...), Q learning etc...
- no-regret learning (Hart, Mas-Colell, Foster, Young, Kalai, Lehrer,...).

In this chapter we will discuss fictitious play.

7.1 Best response and fictitious play

Let $x(t)$ and $y(t)$ be the actions (past)play of the two players, and let

$$p(s) = \frac{1}{s}\int_0^s x(u)\,du \text{ and } q(s) = \frac{1}{s}\int_0^t y(u)\,du.$$

So $p(s)$ and $q(t)$ is the average of the past actions. Differentiating this gives

$$\dot{p}(s) = \frac{1}{s}x(s) - \frac{1}{s}p(s) \text{ and } \dot{q}(s) = \frac{1}{s}y(s) - \frac{1}{s}q(s).$$

Now assume that a player decides to always play a best-response action:

$$x(s) \in \mathcal{B}R_A(q(s)) \text{ and } y(s) \in \mathcal{B}R_B(p(s)) \text{ for } s \geq 1.$$

Then we obtain the following differential equation (inclusion)

$$\begin{array}{ll} \dot{p}(s) & \in \frac{1}{s}(\mathcal{B}R_A(q(s)) - p(s)) \\ \dot{q}(s) & \in \frac{1}{s}(\mathcal{B}R_B(p(s)) - q(s)) \end{array} \tag{7.1}$$

which is called the *fictitious play* dynamics. Note that is a non-autonomous differential equation. But it is closely related to an autonomous system, because if we take the time-reparametrisation $s = e^t$, then this gives

$$\begin{array}{ll} \dot{p}(t) & = (\mathcal{B}R_A(q(t)) - p(t)) \\ \dot{q}(t) & = (\mathcal{B}R_B(p(t)) - q(t)). \end{array} \tag{7.2}$$

which is the autonomous best-response dynamics from the previous chapter:

Exercise 7.1. Consider the matrix from the Shapley best response dynamics from Example 6.7.

1. Show that the fictitious play dynamics associated to the same system still has the same orbits.
2. For $\beta = 0$ the best response dynamics has a periodic orbit as in Figure 29. So there is a curve $t \mapsto x(t) = ((p(t), q(t))$ so that $x(t+T) = x(t)$ for all t. For the corresponding fictitious play dynamics, the speed along thid orbit decays. What is the analogous equation to $x(t+T) = x(t)$ for all t?

7.2 The no-regret set

Denote the maximal-payoff functions

$$\bar{A}(q) := \max_{\bar{p}\in\Delta} \bar{p}\cdot Aq \quad \text{and} \quad \bar{B}(p) := \max_{\bar{q}\in\Delta} p\cdot B\bar{q}, \tag{7.3}$$

Let us show that playing fictitious dynamics leads to 'no-regret'.

Assume that players A and B have respectively m and n actions.

Definition. A joint probability distribution $P = (p_{ij})$ over $S := \{1,\dots,m\}\times\{1,\dots,n\}$ is a *coarse correlated equilibrium (CCE)* for the bimatrix game (A,B) if (p_{ij}), $i = 1,\dots,m$ and $j = 1,\dots,n$ is a matrix with all entries ≥ 0 and so that $\sum_{ij} p_{ij} = 1$ (so $P = (p_{ij})$ is a joint probability distribution) and if

$$\sum_{i,j} a_{i'j}p_{ij} \leq \sum_{i,j} a_{ij}p_{ij}$$

and

$$\sum_{i,j} b_{ij'}p_{ij} \leq \sum_{i,j} b_{ij}p_{ij}$$

for all i', j'. The set of CCE is also called the *Hannan set*.

Lemma 7.1. The set of NE's can be throught of a subset of the CCE set in the sense that if (p,q) be a NE then $p_{ij} = p_i q_j$ where $(p_1,\dots,p_n) = p$ and $(q_1,\dots,q_n) = q$ is in the CCE set.

Proof. Since (p,q) is a NE, $p \in \mathcal{BR}_A(q)$ and $q \in \mathcal{BR}_B(q)$ and therefore for all probability vectors $\tilde{p},\tilde{q}$

$$\tilde{p}\cdot Aq \leq p\cdot Aq \text{ and } p\cdot B\tilde{q} \leq p\cdot Bq.$$

In particular,

$$e_{i'}\cdot Aq \leq p\cdot Aq \text{ and } p\cdot Be_{j'} \leq p\cdot Bq$$

for all i', j' and so

$$\sum_j a_{i'j}q_j \leq \sum_{i,j} a_{ij}p_iq_j \text{ and } \sum_i b_{ij'}p_i \leq \sum_{i,j} b_{ij}p_iq_j.$$

Since p, q are probability vectors this implies

$$\sum_{i,j} a_{i'j}p_iq_j \leq \sum_{i,j} a_{ij}p_iq_j \text{ and } \sum_{i,j} b_{ij'}p_iq_j \leq \sum_{i,j} b_{ij}p_iq_j.$$

Since $p_{ij} = p_iq_j$ the required inequalities in the definition of CCE hold. So CCE is a generalisation of the notion of NE, considering all joint probability distributions P rather than just product probability distributions. □

One way of viewing the concept of CCE is in terms of the notion of *regret*. Let us assume that two players are (repeatedly or continuously) playing a bimatrix game (A,B), and let $P(t) = (p_{ij}(t))$ be the empirical joint distribution of their past play through time t, that is, $p_{ij}(t)$ represents the fraction of time of the strategy profile (i,j) along their play through time t. Then $\sum_{i,j} a_{ij}p_{ij}(t)$ and $\sum_{i,j} b_{ij}p_{ij}(t)$ are the players' average payoffs in their play through time t.

For $x \in \mathbb{R}$, let $[x]_+$ denote the positive part of x: $[x]_+ = x$ if $x > 0$, and $[x]_+ = 0$ otherwise. Then the expression

$$\left[\sum_{i,j} a_{i'j}p_{ij}(t) - \sum_{i,j} a_{ij}p_{ij}(t)\right]_+$$

can be interpreted as the regret of the first player from not having played action i' every single time throughout the entire past history of play. It is (the positive part of) the difference between the player A's payoff that she would have received if she always played i' and what player A's actual past was, both given that player B would have played the same way as she did. Similarly,

$$[\sum_{i,j} b_{ij'}p_{ij}(t) - \sum_{i,j} b_{ij}p_{ij}(t)]_+$$

is the regret of the second player from not having played j'. This regret notion is sometimes called *unconditional* or *external regret* to distinguish it from the *internal* or *conditional regret*[3]. In this context the set of CCE can be interpreted as the set of joint probability distributions with no regret (i.e. the regret is ≤ 0).

Exercise 7.2. 1. Discuss the notion of no-regret and the CCE set in your own words. You might want to do this exercise after you have read the definition of the CE set in Section 4.1. (No solution will be provided for this question.)

7.3 Fictitious play converges to the no-regret set CCE

We now show that continuous-time FP converges to a subset of CCE, namely the subset for which equality holds for at least one i', j' in (7.4).

Theorem 7.1. Let $(x(t), y(t))$, $t \geq 1$, be a trajectory of FP dynamics (7.1) and consider the probability distribution $P(t) = (p_{ij}(t))$ via

$$p_{ij}(t) = \frac{1}{t}\int_1^t x_i(s)y_j(s)ds.$$

Here $x_i(s)$ is the i-th component of $x(s)$ where $x(s) \in \mathcal{B}R_A(q(s))$ (and similar for $y_j(s)$). Then matrix $P(t)$ converges to a subset of the set of CCE, namely the set of joint probability distributions $P = (p_{ij})$ over $S^A \times S^B$ such that for all $(i', j') \in S^A \times S^B$

$$\sum_{i,j} a_{i'j}p_{ij} \leq \sum_{i,j} a_{ij}p_{ij} \quad \text{and} \quad \sum_{i,j} b_{ij'}p_{ij} \leq \sum_{i,j} b_{ij}p_{ij}, \tag{7.4}$$

where *equality* holds *for at least one* $(i', j') \in S^A \times S^B$. In other words, FP dynamics asymptotically leads to no regret for both players.

Note that

$$p_{ij}(t) = \frac{1}{t}\int_0^t x_i(s)y_j(s)ds$$

defines a probability matrix $P(t)$ (all elements of the matrix sum up to one). Here $p_{ij}(t)$ can be interpreted as the total time the first player and 2nd player played action (ij) at the same

[3]Conditional regret is the regret from not having played an action i' whenever a certain action i has been played, that is, $[\sum_j a_{i'j}p_{ij} - \sum_j a_{ij}p_{ij}]_+$ for some fixed $i \in S^A$.

time. Indeed, if $\mathcal{B}R_A(q(s))$ is a singleton then since $x(s) \in \mathcal{B}R_A(q(s))$ we have that $x(s) = e_i$ for some i and $x_i(s) = 1$ and $x_{i'}(s) = 0$ for $i' \neq i$. If $s \mapsto \mathcal{B}R_A(q(s)), \mathcal{B}R_B(p(s))$ is only multivalued for a discrete values of s, the above interpretation is correct.

When we say that FP converges to a certain set of joint probability distributions, we mean that $P(t)$ obtained this way converges to this set.

Proof of Theorem 7.1. Let $\bar{A}$ and $\bar{B}$ be defined as

$$\bar{A}(q) := \max_{\bar{p} \in \Delta} \bar{p} \cdot Aq \quad \text{and} \quad \bar{B}(p) := \max_{\bar{q} \in \Delta} p \cdot B\bar{q}.$$

We have that

$$\frac{d\bar{A}(q(t))}{dt} = x \cdot A \frac{dq}{dt} \tag{7.5}$$

whenever $\mathcal{B}R_A(q(t))$ is unique and $x \in \mathcal{B}R_A(q(t))$.

Let us check (7.5) in an example (the proof in the general case goes similarly): $q(t) = \begin{pmatrix} t \\ 1-t \end{pmatrix}$, $A = I$. Then $\bar{A}(q(t)) = \max(t, 1-t)$ and so $\frac{d}{dt}\bar{A}(q(t)) = \frac{d}{dt}\max(t, 1-t)$ is equal to -1 when $t \in (0, 1/2)$ and equal to $+1$ when $t \in (1/, 2)$. So let us consider the $x \cdot A\frac{dq}{dt}$ where $x \in \mathcal{B}R_A(q(t))$. Notice that $\mathcal{B}R_A(q(t)) = e_2$ (resp. e_1) for $t \in [0, 1/2)$ (resp. $t \in (1/2, 1]$ and so $x \cdot A\frac{dq}{dt} = x \cdot A \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is indeed equal to $\frac{d}{dt}\bar{A}(q(t))$ when $x \in \mathcal{B}R_A(q(t))$.

Therefore, since $x(t) \in \mathcal{B}R_A(q(t))$ and $y(t) \in \mathcal{B}R_B(p(t))$ for $t \geq 1$,

$$\frac{d}{dt}\left(t\bar{A}(q(t))\right) = \bar{A}(q(t)) + t\frac{d}{dt}\left(\bar{A}(q(t))\right) = \bar{A}(q(t)) + tx(t) \cdot A\frac{dq(t)}{dt}.$$

Using first the definition of Fictious Play (7.1) and then (7.5) (which implies that $\bar{A}(q(t)) = x(t) \cdot Aq(t)$), it follows that

$$\frac{d}{dt}\left(t\bar{A}(q(t))\right) = \bar{A}(q(t)) + x(t) \cdot A\left(y(t) - q(t)\right) = x(t) \cdot Ay(t)$$

for $t > 1$. Integrating this equation, we conclude that for $T > 1$,

$$\int_1^T x(t) \cdot Ay(t)\, dt = T\bar{A}(q(T)) - \bar{A}(q(1)),$$

and therefore

$$\lim_{T \to \infty} \left(\frac{1}{T} \left(\int_1^T x(t) \cdot Ay(t)\, dt \right) - \bar{A}(q(T)) \right) = 0.$$

Note that

$$\frac{1}{T}\int_1^T x(t) \cdot Ay(t)\, dt = \sum_{i,j} a_{ij} p_{ij}(T),$$

where, as before, $P(T) = (p_{ij}(T))$ is the empirical joint distribution of the two players' play through time T. On the other hand,

$$\bar{A}(q(T)) = \max_{i'} \sum_j a_{i'j} q_j(T) = \max_{i'} \sum_{i,j} a_{i'j} p_{ij}(T).$$

So the last three equations combined gives

$$\lim_{T\to\infty}\left(\sum_{i,j} a_{ij}p_{ij}(T) - \max_{i'}\sum_{i,j} a_{i'j}p_{ij}(T)\right) = 0.$$

By a similar calculation for B, we obtain

$$\lim_{T\to\infty}\left(\sum_{i,j} b_{ij}p_{ij}(T) - \max_{j'}\sum_{i,j} b_{ij'}p_{ij}(T)\right) = 0.$$

It follows that any FP orbit converges to the set of CCE. Moreover, these equalities imply that for a sequence $t_k \to \infty$ so that $p_{ij}(t_k)$ converges, there exist i', j' so that $\sum_{i,j}(a_{ij} - a_{i'j})p_{ij}(t_k) \to 0$ and $\sum_{i,j}(b_{ij} - b_{ij'})p_{ij}(t_k) \to 0$ as $k \to \infty$, proving convergence to the claimed subset (where equality holds for at least one i'). $\square$

Let us denote the average payoffs through time T along an FP orbit as

$$\hat{u}^A(T) = \frac{1}{T}\int_1^T x(t)\cdot Ay(t)\,dt \quad \text{and} \quad \hat{u}^B(T) = \frac{1}{T}\int_1^T x(t)\cdot By(t)\,dt. \tag{7.6}$$

As a corollary to the proof of the previous theorem we get the following

Proposition 7.1. In any bimatrix game, along every orbit of FP dynamics we have

$$\lim_{T\to\infty}\left(\hat{u}^A(T) - \bar{A}(q(T))\right) = \lim_{T\to\infty}\left(\hat{u}^B(T) - \bar{B}(p(T))\right) = 0.$$

where as before

$$\bar{A}(q) := \max_{\bar{p}\in\Delta}\bar{p}Aq \quad \text{and} \quad \bar{B}(p) := \max_{\bar{q}\in\Delta}pB\bar{q},$$

Another consequence of the previous theorem is:

Proposition 7.2. Let (A, B) be a bimatrix game with unique, interior Nash equilibrium (E^A, E^B). If $\bar{A}(q) \geq \bar{A}(E^B)$ and $\bar{B}(p) \geq \bar{B}(E^A)$ for all $(p, q) \in \Delta\times\Delta$, then asymptotically the average payoff along FP orbits is greater than or equal to the Nash equilibrium payoff (for both players).

Of course the payoff depends on the choice of the payoff matrices. The following result (which we shall not prove here) shows that one can always find an equivalent so that the payoff satisfies the assumptions in the previous proposition:

Theorem 7.2. Let (A, B) be an $n \times n$ bimatrix game with unique, interior Nash equilibrium E. Then there exists a linearly equivalent game (A', B'), for which $\bar{A}'(q) > \bar{A}'(E^B)$ and $\bar{B}'(p) > \bar{B}'(E^A)$ for all $p \neq E^A$ and $q \neq E^B$, and so for (A', B') FP payoff Pareto dominates Nash payoff.

Here we way that (A, B) and (A', B') are *linearly equivalent* if and only if

$$\mathcal{B}R_A = \mathcal{B}R_{A'} \text{ and } \mathcal{B}R_B = \mathcal{B}R_{B'}.$$

Notice that $\mathcal{B}R_A = \mathcal{B}R_{A'}$ if A' is obtained by adding a (possible different) multiple of the column vector $\mathbf{1}$ to each of its columns because then there exists a constant c so that $A'q = Aq + c\cdot\mathbf{1}$ for all q.

Exercise 7.3. 1. Of course Proposition (7.3) suggests that you should play FP, rather than Nash in the above game. Discuss what would happen if one of the players starts to deviate from playing FP, and try to preempt the moves of the other player in a more complicated way than through FP. (This is a rather open ended question, and no solution will be provided for this question.)

7.4 FP orbits often give better payoff than Nash

Consider the one-parameter family of 3×3 bimatrix games (A_β, B_β), $\beta \in (0,1)$, given by

$$A_\beta = \begin{pmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{pmatrix}, \qquad B_\beta = \begin{pmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{pmatrix}. \tag{7.7}$$

This family can be viewed as a generalisation of Shapley's game. This system has been shown to give rise to a very rich chaotic dynamics with many unusual and remarkable dynamical features. The game has a unique, completely mixed Nash equilibrium E, where $E = (E^A, E^B)$ where $E^A = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $E^B = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, which yields the respective payoffs

$$u^A(E^B) = \frac{1+\beta}{3} \quad \text{and} \quad u^B(E^A) = \frac{1-\beta}{3}.$$

To check the hypothesis of Proposition 7.2, let $q = (q_1, q_2, q_3)^\top \in \Delta_B$, then

$$\begin{aligned}
\bar{A}(q) &= \max\{q_1 + \beta q_3, q_2 + \beta q_1, q_3 + \beta q_2\} \\
&\geq \frac{1}{3}((q_1 + \beta q_3) + (q_2 + \beta q_1) + (q_3 + \beta q_2)) \\
&= \frac{1}{3}(q_1 + q_2 + q_3)(1+\beta) \\
&= \frac{1+\beta}{3} \\
&= u^A(E^B) = \bar{A}(E^B).
\end{aligned}$$

Moreover, equality holds if and only if

$$q_1 + \beta q_3 = q_2 + \beta q_1 = q_3 + \beta q_2,$$

which is equivalent to $q_1 = q_2 = q_3$, that is, $q = E^B$. We conclude that $\bar{A}(q) > \bar{A}(E^B)$ for all $q \in \Delta_B \setminus \{E^B\}$, and by a similar calculation, $\bar{B}(p) > \bar{B}(E^A)$ for all $p \in \Delta_A \setminus \{E^A\}$. As a corollary to Proposition 7.2 we get the following result.

Theorem 7.3. Consider the one-parameter family of bimatrix games (A_β, B_β) in (7.7) for $\beta \in (0,1)$. Then any (non-stationary) FP orbit Pareto dominates constant Nash equilibrium play in the long run, that is, for all large t we have

$$\hat{u}^A(t) > u^A(E^B) \quad \text{and} \quad \hat{u}^B(t) > u^B(E^A).$$

A conjecture

There are certainly examples of games where the opposite holds, namely where a FP orbit is Pareto dominated by the Nash payoff. However, a numerical study suggests this is extremely rare. For many games FP orbits Pareto dominate Nash play, and conjecturally, for a very large proportion (say %99 percent), FP orbits dominate Nash play for large periods of time.

Exercise 7.4. 1. Consider the matrix from the Shapley best response dynamics from Example 6.7 taking $\beta = 0$. Show that the set of CCE is not a single point. Hint: the the best response dynamics has a periodic orbit as in Figure 29 and use Theorem 7.1.

2. Does the average payoff $\hat{u}^A(t)$ and $\hat{u}^A(t)$ converge as $t \to \infty$ if we are in the Shapley orbit? Hint: use Proposition 7.1.

7.5 Time averages of Replicator Dynamics converge to pseudo-orbits of Fictitious Play

In Lemma 1.5 we saw that the time-average of a one-player RPS game, converges to the periodic orbit of the corresponding Best Response dynamics. It turns out that this relationship holds much more generally:

Theorem 7.4. Let $x(t)$ be a solution of the one-player RD

$$\dot{x}_i = x_i((Ax)_i - x \cdot Ax), i = 1, \dots, n \tag{RD}$$

and let E be the interior Nash equilibrium. Define $X(t) = \frac{1}{t}\int_0^t x(s)\, ds$. Then there exists $\alpha(t)$ with $\alpha(t) \to 0$ as $t \to \infty$ so that

$$\begin{aligned} x(t) &\in BR_A^{\alpha(t)}(X(t)) \quad \text{and} \\ \dot{X}(t) &\in \frac{1}{t}[BR_A^{\alpha(t)}(X(t)) - X(t)]. \end{aligned} \tag{PFP}$$

So the time-average of the solution of a replicator system converges to a pseudo-orbit of FP dynamics.

Similarly, let $(x(s), y(s))$ be the solution of the two-player RD

$$\begin{cases} \dot{x}_i &= x_i((Ay)_i - x \cdot Ay) \\ \dot{y}_j &= y_j((Bx)_j - y \cdot Bx) \end{cases} \quad \forall x, y \in \Delta, i, j = 1, \dots, n, \tag{RD2}$$

and let $(E_A \times E_B) \in \Delta \times \Delta$ be the interior Nash equilibrium. Let $X(t) = \frac{1}{t}\int_0^t x(s)\, ds$ and $Y(t) = \frac{1}{t}\int_0^t y(s)\, ds$. Then

$$\begin{aligned} x(t) &\in BR_A^{\alpha(t)}(Y(t))\,, & y(t) &\in BR_B^{\alpha(t)}(X(t))\,, \\ \dot{X}(t) &\in \frac{1}{t}[BR_A^{\alpha(t)}(Y(t)) - X(t)]\,, & \dot{Y}(t) &\in \frac{1}{t}[BR_A^{\alpha(t)}(X(t)) - Y(t)]. \end{aligned} \tag{7.8}$$

Proof. Let L be the logit function $L\colon \mathbb{R}^n \to \Delta$ defined by

$$L(x) = \left(\frac{\exp(x_1)}{\sum_j \exp x_j}, \dots, \frac{\exp(x_n)}{\sum_j \exp x_j} \right).$$

It follows from this expression that there exists a function $\alpha(t) \to 0$ as $t \to \infty$ so that for each $x \in \Delta$,

$$L(tx) \in BR^{\alpha(t)}(x). \tag{7.9}$$

Here, as before, the set-valued map $\mathcal{B}R^{\alpha(t)}(x)$ is defined so that its graph is equal to the α-neighbourhood of the graph of $x \mapsto BR(x)$, where $\alpha(t)$ tends to zero as $t \to \infty$.

The two player case: Let $(x(t), y(t))$ be a solution of two-player RD where we assume that $x(0) = (x_0^1, \dots, x_0^n) \in \text{int}\,(\Delta)$. Define $U_s = Ay(s)$, $\bar{U}_t = \frac{1}{t}\int_0^t U_s\, ds$, $\hat{U}_0^k = \log x_0^k$, $\hat{U}_0 = (\hat{U}_0^1, \dots, \hat{U}_0^n)$ and $\xi(t) = L(\hat{U}_0 + \int_0^t U_s\, ds)$. Moreover, define $X(t) = \frac{1}{t}\int_0^t x(s)\, ds$ and $Y(t) = \frac{1}{t}\int_0^t y(s)\, ds$. So, since $U_t = Ay(t)$ we obtain

$$\bar{U}_t = AY(t).$$

An explicit calculation, shows that $\xi(0) = x(0)$ and that $\dot{\xi}$ satisfies the first equation in (RD2). Indeed, differentiating $\log \xi(t)$ with respect to t we obtain

$$\begin{aligned} \frac{\dot{\xi}(t)}{\xi(t)} &= \frac{x_0^k U_t^k \exp \int_0^t U_s\, ds}{x_0^k \exp \int_0^t U_s\, ds} - \frac{\sum_j x_0^j U_t^j \exp \int_0^t U_s^j\, ds}{\sum_j x_0^j \exp \int_0^t U_s^j\, ds} \\ &= U_t^k - \sum_j \frac{x_0^j U_t^j \exp \int_0^t U_s^j\, ds}{\sum_m x_0^m \exp \int_0^t U_s^j\, ds} \\ &= U_t^k - \sum_j U_t^j x_t^j = Ay^k(t) - x(t) \cdot Ay(t), \end{aligned}$$

where the penultimate equality uses the definition of $\xi(t)$ via the logit function. It follows $\xi(t)$ and $x(t)$ satisfy the same differential equations, and so by uniqueness of solutions of RD we get $\xi(t) \equiv x(t)$. Hence, $X(t)$ and $Y(t)$ satisfy

$$\begin{cases} \dot{X}(t) &= \frac{1}{t}(x(t) - X(t)), \\ \dot{Y}(t) &= \frac{1}{t}(y(t) - Y(t)). \end{cases}$$

In particular, since $\overline{U}_t = AY(t)$ and using (7.9),

$$\begin{aligned} x(t) &= \xi(t) = L(\hat{U}_0 + \textstyle\int_0^t U_s\, ds) = L(t[\hat{U}_0/t + \overline{U}_t]) \\ &\in BR^{\alpha(t)}(\overline{U}_t) = BR_A^{\alpha(t)}(Y(t)). \end{aligned} \tag{7.10}$$

Interchanging the roles of $x(t)$ and $y(t)$ (and defining $V_s = Bx(s), \hat{V}_0, \bar{V}_t$ as the analogues of $U_s, \hat{U}_0, \bar{U}_t$) we also get

$$\begin{aligned} y(t) &= L(\hat{V}_0 + \textstyle\int_0^t V_s\, ds) = L(t[\hat{V}_0/t + \overline{V}_t]) \\ &\in BR^{\alpha(t)}(\overline{V}_t) = BR_B^{\alpha(t)}(X(t)) \end{aligned} \tag{7.11}$$

where $\alpha(t) \to 0$ as $t \to \infty$ because of (7.9). It follows that for the time averages $X(t), Y(t)$ of the replicator systems one has

$$\begin{cases} \dot{X}(t) &\in \frac{1}{t}(BR_A^{\alpha(t)}(Y(t)) - X(t)) \\ \dot{Y}(t) &\in \frac{1}{t}(BR_B^{\alpha(t)}(X(t)) - Y(t)) \end{cases} \tag{7.12}$$

where again $\alpha(t) \to 0$ as $t \to \infty$. Note that $Y \mapsto BR_A^{\alpha}(Y)$ is a neighbourhood of the (set-valued) graph of $Y \mapsto BR_A(Y)$, and whenever $BR_A(Y)$ is single valued we have $BR_A^{\alpha}(Y) \to BR_A(Y)$ as $\alpha \to 0$.

The one-player case. Let $x(t)$ be a solution of the equation $\dot{x}_i = x_i((Ax)_i - x' cdot Ax)$. Define $U_s = Ax(s)$, $\bar{U}_t = \frac{1}{t}\int_0^t U_s\,ds$, $\hat{U}_0^k = \log x_0^k$, and $\xi(t) = L(\hat{U}_0 + \int_0^t U_s\,ds)$ and $X(t) = \frac{1}{t}\int_0^t \xi(s)\,ds$. The same calculation as before gives $\xi(0) = x(0)$ and that $\xi(t)$ and $x(t)$ satisfy the same differential equation. Hence $\xi(t) = x(t)$ for all t. It follows as before that

$$\begin{array}{rl} X(t) & = \frac{1}{t}\int_0^t x(s)\,ds, \\ x(t) & = L(\hat{U}_0 + tAX(t)) \,\in BR_A^{\alpha(t)}(X(t)), \\ \dot{X}(t) & \in \dfrac{1}{t}[BR_A^{\alpha(t)}(X(t)) - X(t)]. \end{array} \tag{7.13}$$

□

7.6 Discrete fictitious dynamics

Sometimes it is more natural to consider discrete time, so assume that $t \in \mathbb{N}$. In this case we let $p(0), q(0)$ be the a priori believe at time $t = 0$ of the probability that player B resp A thinks the strategies will be played. The updating rule about these believes is then

$$p(n+1) = \frac{np(n) + e_i(n)}{n+1}, q(n+1) = \frac{nq(n) + e_j(n)}{n+1}$$

where

$$e_i(n) \in \mathcal{B}R_A(q_n) \text{ and } e_j(n) \in \mathcal{B}R_B(p_n).$$

So

$$p(n+1) - p(n) = \frac{1}{n}(e_i(n) - p(n)), q(n+1) = q(n) + \frac{1}{n}(e_j(n) - q(n)).$$

This should be considered as the discrete approximation of the continuous best response dynamics

$$\dot{p} = \mathcal{B}R_A(q) - p, \dot{q} = \mathcal{B}R_B(p) - q.$$

Exercise 7.5. Write computer code which draws orbits of the discrete fictitious play dynamics associated to the game corresponding to matrices (7.7) (taking various choices for β including $\beta = 0$, $\beta =$ the golden mean and $\beta = 0.9$).

8 Conclusion

Let us review what we saw in these lecture notes.

8.1 Relationship between all these learning mechanisms

The learning algorithms described above originate from entirely disjoint communities, but although seemingly quite different they are surprisingly connected:

$$\begin{tikzcd}
\text{Fictitious Play (FP)} \arrow[rr, leftrightarrow, "B"] \arrow[dr, leftrightarrow, "A"'] & & \text{Reinforcement Learning (RL)} \arrow[dl, leftrightarrow, "C"] \\
& \text{Replicator Dynamics (RD)} &
\end{tikzcd}$$

A The time average of a replictor orbit corresponds to a pseudo-orbit of fictitious play dynamics, see Section 5.5.
Vice versa, if there $\exists$ hyperbolic orbits of FP $\implies$ $\exists$ corresponding orbit in (RD) Castro & SvS - in preparation. In particular, as we know there is chaotic dynamics in FP (SvS & Sparrow), one also can rigorously show that there exists chaotic switching in (RD).

B Reinforcement learning with choosing ϵ-greedy choices is very closely related to the type of dynamics one sees in Best Response dynamics and Fictitious Play. This connection is explored in work by SvS & Winckler - in preparation, see also Wunder, Littman, Babes.

C was known since 90's e.g. Börgers. Revisited by e.g. Sato, Akiyama and Crutchfield and also Tuyls et.al.

On the other hand, at this moment, it seems not so clear what the connections are of these learning algorithms with No-Regret learning. (Best response dynamics converges to the CCE set, whereas the no-regret learning algorithm converges to the somewhat smaller CE set.)

That there are such relationships is a little surprising as the underlying mechanisms and approaches are somewhat different:

- replicator dynamics (RD) encourages 'fitness',
- fictitious play (FP) keeps tracks of the average of the other player's actions and gives a best response to that, whereas
- reinforcement learning (RL) keeps track of past payoff and responds to that.

8.2 Quite often these learning mechanisms lead to complicated dynamics

This is rigorously proved for Best Response dynamics for the family of 3×3 Rock-Paper-Scissor games discussed in Example 6.7 by SvS and Sparrow. This was also shown numerically for the replicator dynamics by Sato, Akiyama and Crutchfield.

Several classes of learning dynamics was considered by Galla & Farmer and they found that for many games these lead to complicated or even chaotic dynamics.

8.3 Complicated dynamics quite often leads to better payoff performance

One might think that any behaviour away from the NE is bad for the players. This is of course not true. One simple instance which shows this in the coordination games in Exercise 4.1.2.

This point of view is taken much further in Ostrovski & SvS where they show that the average payoff for **both** players is very often better if they play (FP) than if they play (NE). It would be interesting to explore whether this is also true for the other learning dynamics considered in these lecture notes (or specifically for the systems considered by Galla & Farmer).

A Appendix

A.1 Existence and uniqueness of solutions of ODE

Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a C^1 function.

1. For each $x_0 \in \mathbb{R}^n$ there exists $\epsilon > 0$ so that the initial value problem

$$\dot{x} = f(x), x(0) = x_0 \tag{A.1}$$

   has a *unique solution* $x\colon (-\epsilon, \epsilon) \to \mathbb{R}^n$ with $x(0) = x_0$.

2. The solution of (A.1) can be extended to a *maximal solution*: there exist $a(x_0) < -\epsilon < \epsilon < b(x_0)$ and a function $(a(x_0), b(x_0)) \ni t \mapsto x(t) \in \mathbb{R}^n$ which satisfies (A.1). The interval $(a(x_0), b(x_0))$ is maximal in the sense that if $|a(x_0)| < \infty$ then $|x(t)| \to \infty$ as $t \downarrow a(x_0)$ and similarly if $|b(x_0)| < \infty$ then $|x(t)| \to \infty$ as $t \uparrow b(x_0)$.

A.2 Some further background on ODE's

1. For the purposes of this notes we say that $V\colon \mathbb{R}^n \to \mathbb{R}$ is a *Lyapounov function* if

$$\frac{dV(x(t))}{dt} < 0.$$

   Often this derivative is denoted by $\dot{V}$. The fact that $\dot{V} < 0$ implies that V decreases along solutions. Quite often a Lyapounov function are also assumed to achieve a minimum in some point $\bar{x}$ and then this can be used to show that $x(t) \to \bar{x}$ as $t \to \infty$.

2. A point $\bar{x}$ is called *Lyapounov stable* if for each $\epsilon > 0$ there exists $\delta > 0$ so that if $|x(0) - \bar{x}| < \delta$ then $|x(t) - \bar{x}| < \epsilon$ for all $t \geq 0$.

3. A point $\bar{x}$ is called *asymptotically stable* if there exists $\delta > 0$ so that if $|x(0) - \bar{x}| < \delta$ then $x(t) \to \bar{x}$ as $t \to \infty$.

4. A well-known theorem states the following: Assume that $U \subset \mathbb{R}^n$ is an open set, $\bar{x} \in U$ and $V\colon U \to \mathbb{R}$ is a function so that $V(x) > 0$ for all $x \neq \bar{x}$ and $V(\bar{x}) = 0$. Then

   (a) If $\dot{V} \leq 0$ then $\bar{x}$ is stable.

   (b) If $\dot{V} < 0$ for $x \in U \setminus \{\bar{x}\}$ then $\bar{x}$ is asymptotically stable.

5. The *omega-limit set* of a point x is defined as follows. Let $x(t)$ be the solution of the ODE with $x(0) = x$. Then

$$\omega(x) := \{y; x(t_k) \to y \text{ for some sequence } t_k \to \infty\}.$$

A.3 Stable and unstable manifolds at singularities of vector fields

A point $\bar{x}$ so that $f(\bar{x}) = 0$ is called a *singularity* of the vector field $f\colon \mathbb{R}^n \to \mathbb{R}^n$. This singularity is called *hyperbolic* if all eigenvalues of the Jacobian matrix $B = (Df)_{\bar{x}}$ are off the imaginary axis. We then have the following. The sets

$$W^s(\bar{x}) = \{x; \phi_t(x) \to \bar{x} \text{ as } t \to \infty\}$$

$$W^u(\bar{x}) = \{x; \phi_t(x) \to \bar{x} \text{ as } t \to -\infty\}$$

are (immersed) manifolds which are tangent to $\bar{x}$ to the eigenspace of B associated to the eigenvalues with negative real part respectively positive real part. Here $\phi_t(x)$ is the solution of $\dot{x} = f(x)$ so that $\phi_0(x) = x$. More precisely

$$\frac{d\phi_t(x)}{dt} = f(\phi_t(x)), \phi_0(x) = x.$$

Usually $\phi_t(x)$ is called the flow of the differential equation (or of the vector field).

A.4 Chain recurrence and attractors

Let Φ_t be a flow on a manifold M.

Definition. Let $\delta, T > 0$ and $a, b \in M$. A (δ, T)-*pseudo-orbit* from a to b is a finite sequence of points $a = x_0, x_1, \ldots, x_n = b \in M$ such that there exist $t_0, \ldots, t_{n-1} > T$ with $d(\Phi_{t_i}(x_i), x_{i+1}) < \delta$ for $i = 0, \ldots, n-1$.

In a pseudo-orbit we do not need to follow the same trajectory forever, but we are allowed to jump a finite number of times, and the parameters δ, T control the maximum size of the jump and the minimum duration between two subsequent jumps. Letting $\delta \to 0$ and $T \to \infty$ leads us to the concept of chain recurrence and chain transitivity.

Definition. A point $x \in M$ is *chain recurrent*, if for every choice of $\delta, T > 0$ there is a (δ, T)-pseudo-orbit from x to itself. We denote by $R(\Phi)$ the set of all chain recurrent points of the flow Φ. We say that Φ is chain recurrent if $R(\Phi) = M$.

Definition. The flow Φ is called *chain transitive*, if for every pair $a, b \in M$ and every choice of $\delta, T > 0$ there is a (δ, T)-pseudo-orbit from a to b.

Definition. A compact subset $A \subset M$ is called invariant, if $\Phi_t(A) = A$ for all $t > 0$. An invariant subset is called *internally chain recurrent (transitive)*, if the restricted flow $\Phi_t|A$ is chain recurrent (transitive).

Definition. A compact, invariant subset $A \subset M$ is called an *attractor*, if $A \neq \emptyset$ and if there exists an open set O with $A \subset O$ and $dist(\Phi_t(O), A) \to 0$ as $t \to \infty$. It is called a proper attractor if, in addition, $A \subset M$.

A.5 Convex sets and functions

A set $\mathcal{C} \subset \mathbb{R}^k$ is convex if for each $x, y \in \mathcal{C}$ and each $\lambda \in [0, 1]$ one has that $\lambda x + (1-\lambda)y \in \mathcal{C}$.

A function $f : \mathcal{C} \to \mathbb{R}$ is called *convex* if for each $x, y \in \mathcal{C}$ and each $\lambda \in [0, 1]$ we have that $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Similarly, f is *concave* if the reverse inequality holds.

A.6 The origins of Q-learning

Q-learning has its roots in the field of Markov decision processes (MDP). Let us give a short summary of this, and explain why the assumption that there is a stationary distribution is violated in a game-theoretic setting. Assume that there are a finite number of states $s \in S$ and a finite number of actions. Now assume that there the probability of moving from state s to s' is determined by the entry $P_{s,s'}$ of a probability matrix P.

Let us first assume there is only one action. In this case (MDP) reduces to Markov reward processes (MRP). So assume you get a reward R_t at time t, and discount the rewards in the future by a factor $\gamma \in (0,1)$. Then your wealth at time t is defined to be equal to $G_t = \sum_{k\geq 0} \gamma^k R_{t+k+1}$ where $R_t, R_{t+1}, \ldots$ are i.i.d. random variables. The value of being in state $s \in S$ is then defined to be

$$v(s) = \mathbb{E}(G_t|S_t = s) = \mathbb{E}\left(R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots |S_t = s\right)$$
$$= \mathbb{E}(R_{t+1} + \gamma G_{t+1}|S_t = s) = \mathbb{E}(R_{t+1}|S_t = s) + \gamma \sum_{s' \in S} P_{s,s'} \nu(s')).$$

which reduces in short to

$$v(s) = R_s + \gamma \sum_{s' \in S} P_{s,s'} \nu(s'))$$

and so in factor form

$$v = R + \gamma P v.$$

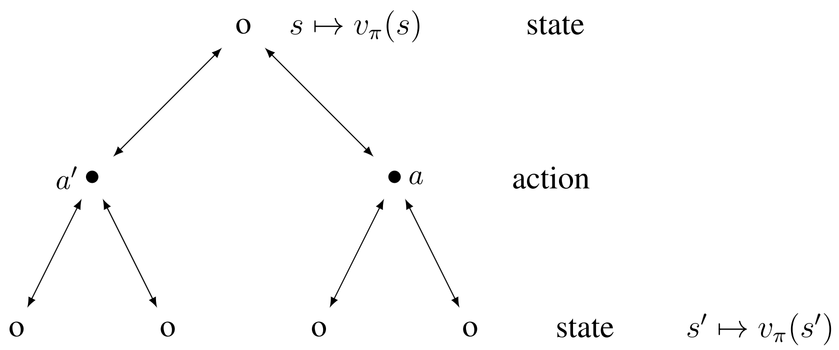
In this case the value v can be computed from this equation.

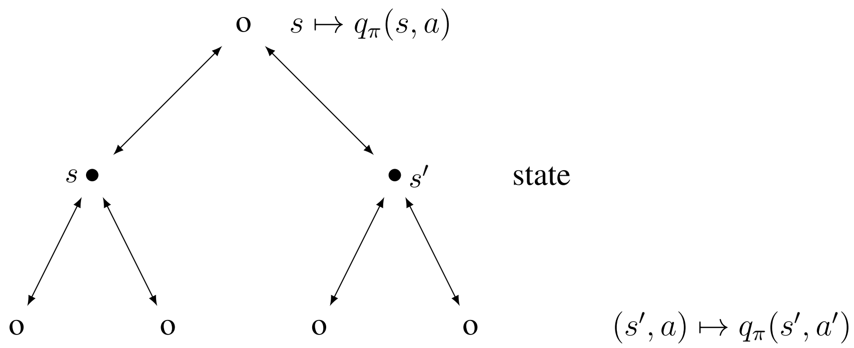
Now assume that there are several actions, so that P^a and R^a_s depend on which action a was chosen:

$$P^a_{s,s'} = \mathbb{P}(S_{t+1} = s'|S_t = s, A_t = a), R^a_s = \mathbb{E}(R_{t+1}|S_t = s, A_t = a).$$

A policy (or strategy) π is an assignment of probabilities of actions for each state:

$$\pi(a|s) = \mathbb{P}(a_t = a|S_t = s).$$





We can define the state value function

$$v_\pi(s) = \mathbb{E}_\pi(G_t|S_t = s) = \sum_{a\in A}\left(R^a_s + \gamma \sum_{s'\in S} P_{s,s'}\nu(s')\right).$$

and the state-action value function

$$q_\pi(s,a) = \mathbb{E}_\pi(G_t|S_t = s, A_t = a) = R^a_s + \gamma \sum_{s'\in S} P_{s,s'}\left(\sum_{a'\in A}\pi(a'|s')q_\pi(s',a')\right).$$

We say that π is an optimal policy if $\nu_\pi(s) \geq \nu_{\pi'}(s)$ for each other policy π'.

Theorem A.1. For each (MDP) there exists a not necessarily unique optimal policy π_* and

$$v_{\pi_*} = \max_{\pi'} \nu_{\pi'} \text{ and } q_{\pi_*}(s,a) = \max_{\pi'} q_{\pi'}(\pi, a).$$

Given q_* find π_* by $\pi_*(a|s) = 1$ is $a = \arg\max q_*(s,a)$. We have the Bellman optimality equation:

$$v_*(s) = \max_a q_*(s,a)$$

$$q_*(s,a) = R^a_s + \gamma \sum_{s'\in S} P^a_{\pi'} v_*(s)$$

and therefore

$$q_*(s,a) = R^a_s + \gamma \sum_{s'\in S} P^a_{\pi'} \max_{a'} q_*(s',a')$$

Note that an important assumption in this theorem is that the distributions are stationary. This assumption is violated in a game theoretic setting.

Q learning, which we discussed in the last section of Chapter 6, is aimed at numerically finding q_*.

B Python Code

Below we include some python code which we've used during the course.

B.1 Code for computing orbits of one-player replicator dynamics with three strategies

replicator-3x1

October 7, 2020

[41]:
```
# Import the required modules
import numpy as np
import matplotlib.pyplot as plt
# This makes the plots appear inside the notebook
%matplotlib inline
from scipy.integrate import odeint
```

0.0.1 Solving replicator equations with python

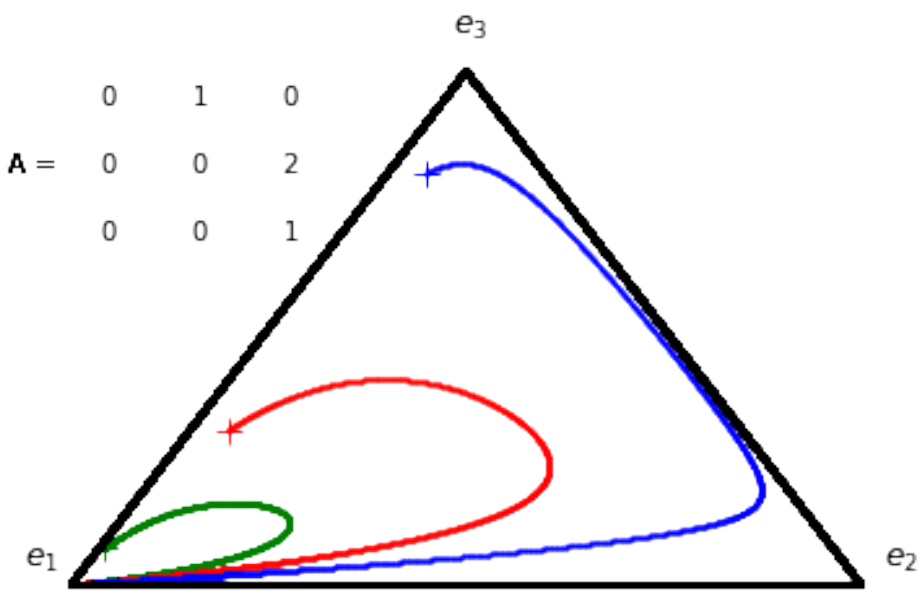
$\frac{dx_i}{dt} = x_i[(Ax)_i - x'Ax]$

[52]:
```
# define a projection from the 3D simplex on a triangle
proj = np.array(
    [[-1 * np.cos(30. / 360. * 2. * np.pi),np.cos(30. / 360. * 2. * np.pi),0.],
     [-1 * np.sin(30. / 360. * 2. * np.pi),-1 * np.sin(30. / 360. * 2. * np.
 ↪pi),1.]])
# project the boundary on the simplex onto the boundary of the triangle
ts = np.linspace(0, 1, 10000)
PBd1 = proj@np.array([ts,(1-ts),0*ts])
PBd2 = proj@np.array([0*ts,ts,(1-ts)])
PBd3 = proj@np.array([ts,0*ts,(1-ts)])
```

[60]:
```
# choose game
# game Ex 1.7 notes
A = np.array([[ 0, 1 ,0], [ 0, 0 ,2], [ 0, 0 ,1]]) # row, 2nd row, 3rd row
x01 = np.array([0.92, 0.01, 0.07])
x02 = np.array([0.65,0.05,0.3])
x03 = np.array([0.15, 0.05, 0.8])

#define replicator equation
def replicator(x,t):
    return x * (A@x - np.transpose(x) @ (A@x))

# compute orbits
ts = np.linspace(0,100,10000)
xt1 = odeint(replicator, x01, ts)
```

```
xt2 = odeint(replicator, x02, ts)
xt3 = odeint(replicator, x03, ts)
```

[62]:
```
# project the orbits on the triangle
orbittriangle1=proj@xt1.T
orbittriangle2=proj@xt2.T
orbittriangle3=proj@xt3.T
ic1=proj@x01
ic2=proj@x02
ic3=proj@x03

# no box
plt.box(False)
plt.axis(False)

# plot the orbits, the initial values, the corner points,  and the boundary␣
↪points
plt.plot(orbittriangle1[0],orbittriangle1[1],".",markersize=1,color='green')
plt.plot(orbittriangle2[0],orbittriangle2[1],".",markersize=1,color='red')
plt.plot(orbittriangle3[0],orbittriangle3[1],".",markersize=1,color='blue')

plt.plot(ic1[0],ic1[1],"+",markersize=10,color='green')
plt.plot(ic2[0],ic2[1],"+",markersize=10,color='red')
plt.plot(ic3[0],ic3[1],"+",markersize=10,color='blue')

plt.text(-0.8660254-0.1, -0.5 +0.05 , "$e_1$",fontsize=12)
plt.text(+0.8660254+0.05, -0.5 +0.05 , "$e_2$",fontsize=12)
plt.text(0-0.03, 1 +0.1 , "$e_3$",fontsize=12)

plt.plot(PBd1[0], PBd1[1], ".",color='black',markersize=3)
plt.plot(PBd2[0], PBd2[1], ".",color='black',markersize=3)
plt.plot(PBd3[0], PBd3[1], ".",color='black',markersize=3)

# add the game matrix in the figure
for i in [0,1,2]:
    for j in [0,1,2]:
        c = A[i][j]
        plt.text(0.2*j-0.8, -0.2*i+0.9, str(c))
        plt.text(0.3-1.3,0.7,"A =")
#plt.text(0-0.03, 1 +0.1 ,A[0,0],A[0,1],A[0,2] ,fontsize=12)


#plt.plot(pE[0],pE[1],"+")
plt.savefig("Plots/flowportrait.pdf")
```


e_3
$\mathbf{A} = \begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{matrix}$
e_1
e_2

[ ]:

B.2 Code for time averages of RPS 1-player

replicator-RPS-1player-timeaverag

November 1, 2020

[1]:
```
# Import the required modules
import numpy as np
import matplotlib.pyplot as plt
# This makes the plots appear inside the notebook
%matplotlib inline
from scipy.integrate import odeint, solve_ivp
import copy
```

0.0.1 Solving replicator equations with python

$\frac{dx_i}{dt} = x_i[(Ax)_i - x'Ax]$ here we take the RPS game and also compute the corresponding expression $average(T) = \frac{1}{T}\int_0^T x(s)ds$.

[2]:
```
# define a projection from the 3D simplex on a triangle
proj = np.array(
    [[-1 * np.cos(30. / 360. * 2. * np.pi),np.cos(30. / 360. * 2. * np.pi),0.],
     [-1 * np.sin(30. / 360. * 2. * np.pi),-1 * np.sin(30. / 360. * 2. * np.
 ↪pi),1.]])
# project the boundary on the simplex onto the boundary of the triangle
ts = np.linspace(0, 1, 10000)
PBd1 = proj@np.array([ts,(1-ts),0*ts])
PBd2 = proj@np.array([0*ts,ts,(1-ts)])
PBd3 = proj@np.array([ts,0*ts,(1-ts)])
```

[3]:
```
# choose game
# game Ex 1.7 notes
# Rock Paper Scissors
b=2
A = np.array([[ 0, 1 ,-b], [ -b, 0 ,1], [ 1, -b ,0]]) # row, 2nd row, 3rd row
x01 = np.array([0.3, 0.2, 0.5])

#define replicator equation
def replicator(x,t):
    return x * (A@x - np.transpose(x) @ (A@x))

# end time and spacing of observed times
```

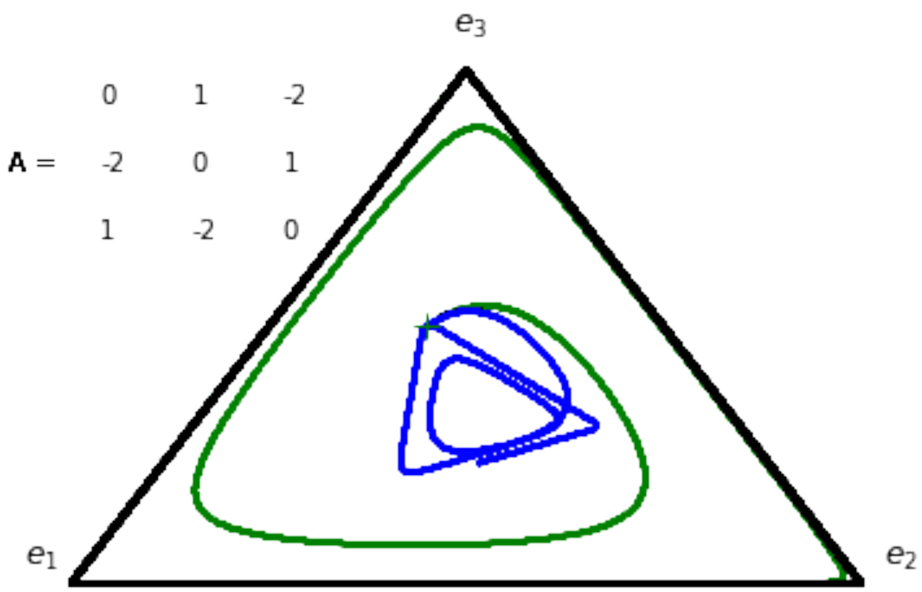
```python
endtime=100
steps=1000000

# compute orbits
ts = np.linspace(0,endtime,steps)
xt1 = odeint(replicator, x01, ts)

#t_step = 0.03
#t_final =100
#time = np.arange(0,t_final,t_step)
#sol = solve_ivp(replicator, [0,t_final], x01, method='DOP853', t_eval=time)
```

[4]:
```python
# compute average of orbit
average1=copy.deepcopy(xt1)
r=endtime/steps
for n in range(1, steps-1):
    average1[n]= (n/(n+1)) * average1[n-1] + (1/(n+1)) * xt1[n]
    # project average along orbit also in the triangle
average_proj=proj@average1.T

SumVector=np.cumsum(xt1[1:steps],axis=0)
print(SumVector.shape)
tt=ts[1:steps]
print(tt.shape)
average2 = SumVector / tt.reshape((steps-1,1))
average2bis = average2/50
average2_proj=proj@average2bis.T
```

```
(999999, 3)
(999999,)
```

[5]:
```python
# project the orbits on the triangle
orbittriangle1=proj@xt1.T
ic1=proj@x01

# no box
plt.box(False)
plt.axis(False)

# plot the orbits, the initial values, the corner points,  and the boundary␣
↪points
plt.plot(orbittriangle1[0],orbittriangle1[1],".",markersize=1,color='green')

# plot the average
# plt.plot(orbittriangle1[0],orbittriangle1[1],".",markersize=1,color='red')
# plt.plot(average[0],average[1],".",markersize=1,color='red')
plt.plot(average_proj[0],average_proj[1],".",markersize=1,color='blue')
```

```
#plt.plot(average2_proj[0],average2_proj[1],".",markersize=1,color='red')

plt.plot(ic1[0],ic1[1],"+",markersize=10,color='green')

plt.text(-0.8660254-0.1, -0.5 +0.05 , "$e_1$",fontsize=12)
plt.text(+0.8660254+0.05, -0.5 +0.05 , "$e_2$",fontsize=12)
plt.text(0-0.03, 1 +0.1 , "$e_3$",fontsize=12)

plt.plot(PBd1[0], PBd1[1], ".",color='black',markersize=3)
plt.plot(PBd2[0], PBd2[1], ".",color='black',markersize=3)
plt.plot(PBd3[0], PBd3[1], ".",color='black',markersize=3)

# add the game matrix in the figure
for i in [0,1,2]:
    for j in [0,1,2]:
        c = A[i][j]
        plt.text(0.2*j-0.8, -0.2*i+0.9, str(c))
        plt.text(0.3-1.3,0.7,"A =")
#plt.text(0-0.03, 1 +0.1 ,A[0,0],A[0,1],A[0,2] ,fontsize=12)


#plt.plot(pE[0],pE[1],"+")
plt.savefig("Plots/flowportrait.pdf")
```



The green curve gives the solution of the replicator system. The blue one is an attempt to compute $1/T \int_0^T x(s)ds$ as a function of T. This is done by "updating the average". This seems to give the correct answer. However, when increasing the endtime something goes wrong. Is this because the solution near the singularity is not too accurate? clearly the cumsum solution (drawn in red) is incorrect.

```
[ ]:
```

```
[ ]:
```

B.3 Python code for computing orbits of two player RPS game

replicatorRPS-Sato

October 7, 2020

[232]:

```
import numpy as np
import matplotlib.pyplot as plt
#import networkx as nx
from scipy.integrate import odeint, solve_ivp, ode
from mpl_toolkits.mplot3d import Axes3D
from scipy import sparse
import time as tm
from scipy.spatial.distance import pdist, squareform
from mpl_toolkits.axes_grid1 import make_axes_locatable
```

3x3 replicator dynamics $\dot{x}_i = x_i(Ay - xAy)$ $\dot{y}_i = y_i(x'B - xBy)$ here we use the 2nd notation

[233]:

```
def replicator(t,x,A,B):
    dx = np.zeros(6)
    dx[:3] = x[:3] * (A@x[3:] - np.transpose(x[:3]) @ (A@x[3:]))
    dx[3:] = x[3:] * (B.T@x[:3] -np.transpose(x[:3]) @ (B@x[3:]))
    return dx
```

[305]:

```
# initial value
z0 = np.random.uniform(0.1, 0.2, 6).T
z0= [0.26, 0.113333, 0.626667, 0.165, 0.772549, 0.062451]
#z0=  [0.05, 0.35, 0.6, 0.1, 0.2, 0.7]

# time interval
t_step = 0.03
t_final =1000
time = np.arange(0,t_final,t_step)
init_time = tm.time()
# # %% Define time spans, initial values, and constants
# tspan = np.linspace(0, 15, 5000)
# ttt=tspan[-1]
# yinit = [-1]

# Define matrices
# Sato's matrices: Sato use 1st notation so take transpose
epsilonx=0
epsilony=-epsilonx
A= np.array([[ epsilonx, 1 , -1], [ -1, epsilonx ,1], [ 1, -1 ,epsilonx]])
```

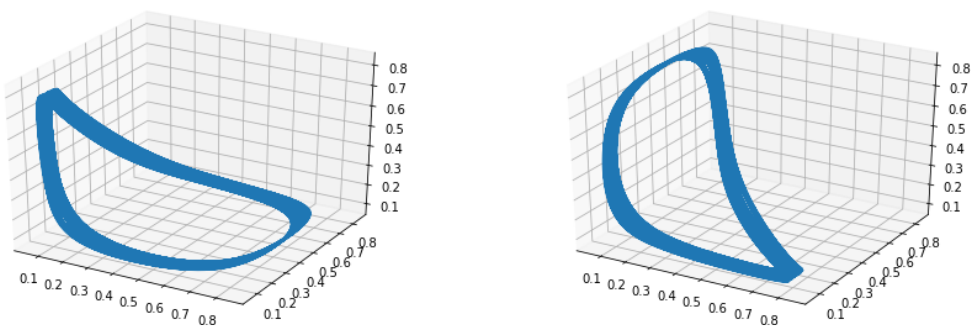
```
BSato= np.array([[ epsilony, 1 ,-1], [ -1, epsilony ,1], [ 1, -1 ,epsilony]])
B=BSato.T
print('A=',A)
print('BSato=',BSato)
print('B=',B)

# integrate ODE
sol = solve_ivp(replicator, [0,t_final], y0=z0, method='DOP853', t_eval=time,␣
↪args=(A,B))

xx = sol.y
xx.shape
```

```
A= [[ 0  1 -1]
 [-1  0  1]
 [ 1 -1  0]]
BSato= [[ 0  1 -1]
 [-1  0  1]
 [ 1 -1  0]]
B= [[ 0 -1  1]
 [ 1  0 -1]
 [-1  1  0]]
```

[305]: (6, 33334)

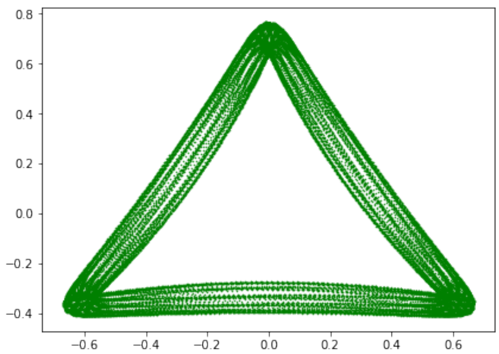
[306]:
```
transient = 100
# plt.plot(xx[1,:])
plt.figure(figsize=(15,5))
plt.subplot(121, projection='3d')
plt.plot(xx[1,transient:], xx[3,transient:], xx[4,transient:])
plt.subplot(122, projection='3d')
plt.plot(xx[1,transient:], xx[2,transient:], xx[4,transient:])
```

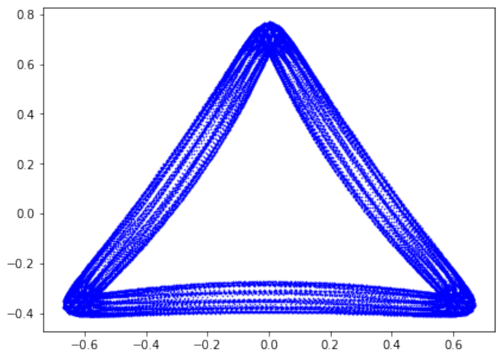
[306]: [<mpl_toolkits.mplot3d.art3d.Line3D at 0x7fbeffcd2c90>]



[307]:
```
# define the projection to triangular coordinates
proj = np.array(
    [[-1 * np.cos(30. / 360. * 2. * np.pi),np.cos(30. / 360. * 2. * np.pi),0.],
     [-1 * np.sin(30. / 360. * 2. * np.pi),-1 * np.sin(30. / 360. * 2. * np.
 ↪pi),1.]])
playerA=xx[0,transient:], xx[1,transient:], xx[2,transient:]
playerB=xx[3,transient:], xx[4,transient:], xx[5,transient:]
orbittriangle1=proj@playerA
orbittriangle2=proj@playerB
plt.figure(figsize=(15,5))
plt.subplot(121)
plt.plot(orbittriangle1[0],orbittriangle1[1],".",markersize=1,color='green')
plt.subplot(122)
plt.plot(orbittriangle2[0],orbittriangle2[1],".",markersize=1,color='blue')
```

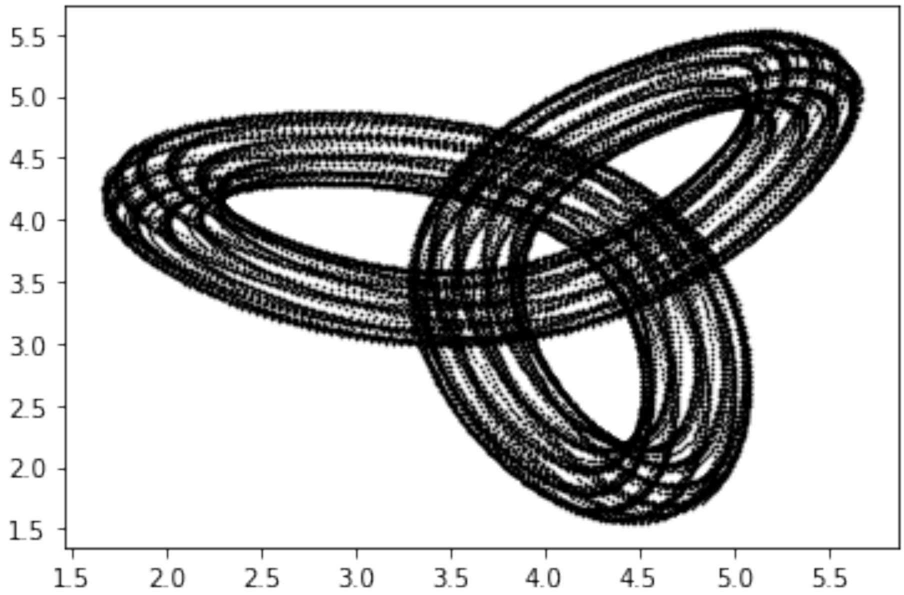
[307]: [<matplotlib.lines.Line2D at 0x7fbf023e4610>]





[308]:
```
PROJ4D2D= np.array([[3.650,-1.350,1.35,5.35,1.35,1.4500],[0.4,0.4,4.6,1.9,-0.
 ↪4,4.4]])
XY= PROJ4D2D @ xx[:,:]
plt.plot(XY[0],XY[1],".",markersize=1,color='black')
```

[308]: [<matplotlib.lines.Line2D at 0x7fbf0297e110>]


5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

B.4 Python code for Exercise 6.1

exercise6p1

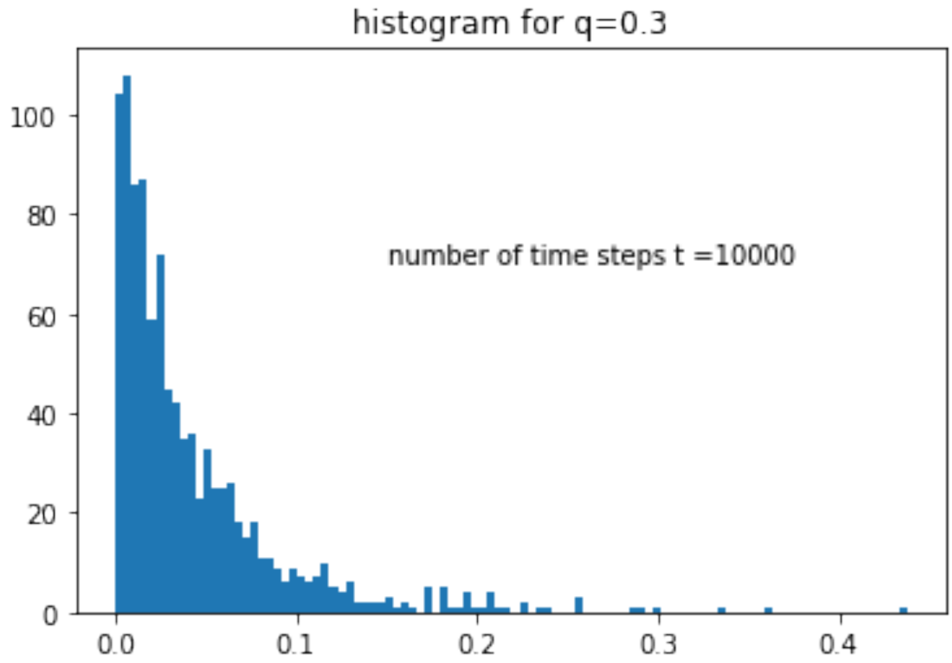
December 23, 2020

[62]:
```
# Import the required modules
import numpy as np
import matplotlib.pyplot as plt
# This makes the plots appear inside the notebook
%matplotlib inline
import random
```

[75]:
```
def flip(p):
    return 1 if random.random() < p else 0
```

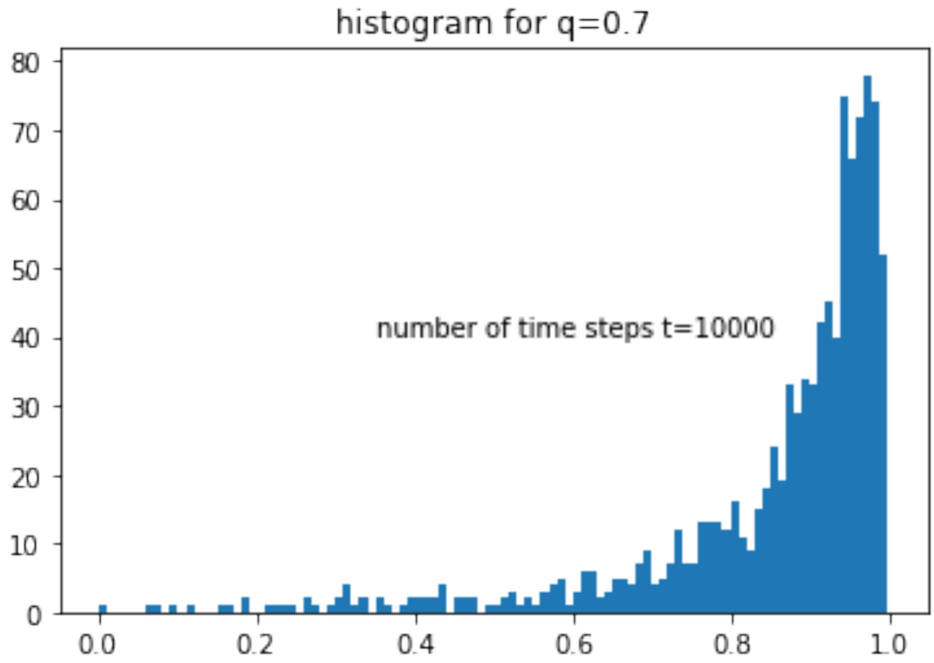
[235]:
```
theta1=10
theta2=10
q=0.3
n=1000 # number of sample path
t=10000 # number of time steps
x=np.zeros(n)
for i in range(0, n):
    f11=0
    f2=0
    for j in range(0, t):
        p1=(theta1+10*f11)/(theta1+theta2+10*f11+5*f2)
        TypeI=flip(q)
        Med=flip(p1)
        f11+=Med*TypeI
        f2+=(1-Med)
    x[i]=p1
```

[237]:
```
plt.hist(x, bins = 100)
plt.title("histogram for q="+ str(q))
plt.text(q/2, 70, "number of time steps t ="+str(t))
plt.show()
```



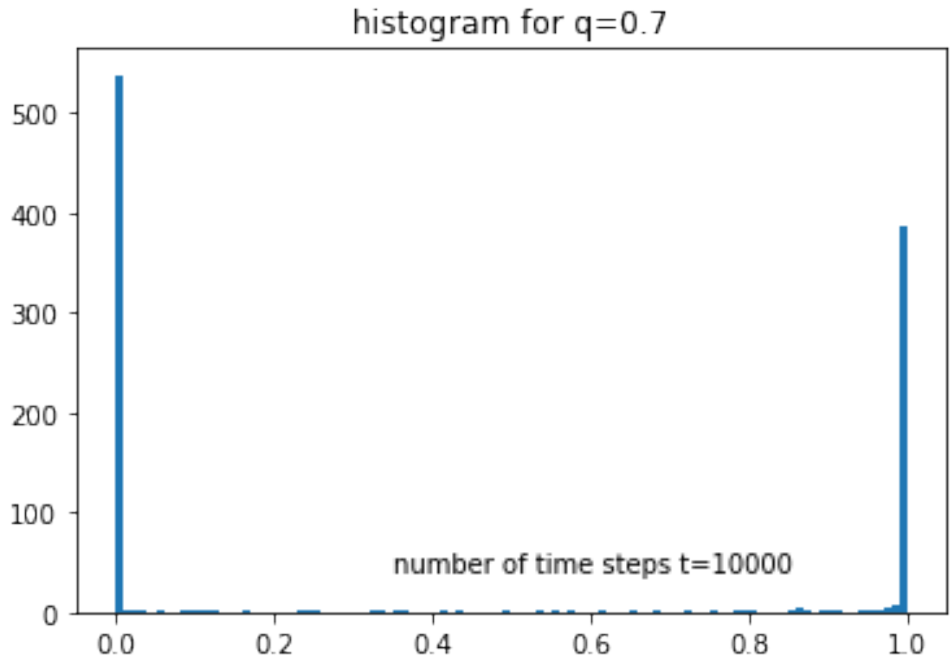
[262]:
```
theta1=10
theta2=10
q=0.7
n=1000 # number of sample path
m=10000 # number of time steps
x=np.zeros(n)
for i in range(0, n):
    f11=0
    f2=0
    for j in range(0, t):
        p1=(theta1+10*f11)/(theta1+theta2+10*f11+5*f2)
        TypeI=flip(q)
        Med=flip(p1)
        f11+=Med*TypeI
        f2+=(1-Med)
    x[i]=p1
```

[246]:
```
plt.hist(x, bins = 100)
plt.title("histogram for q="+ str(q))
plt.text(q/2, 40, "number of time steps t="+str(t))
plt.show()
```



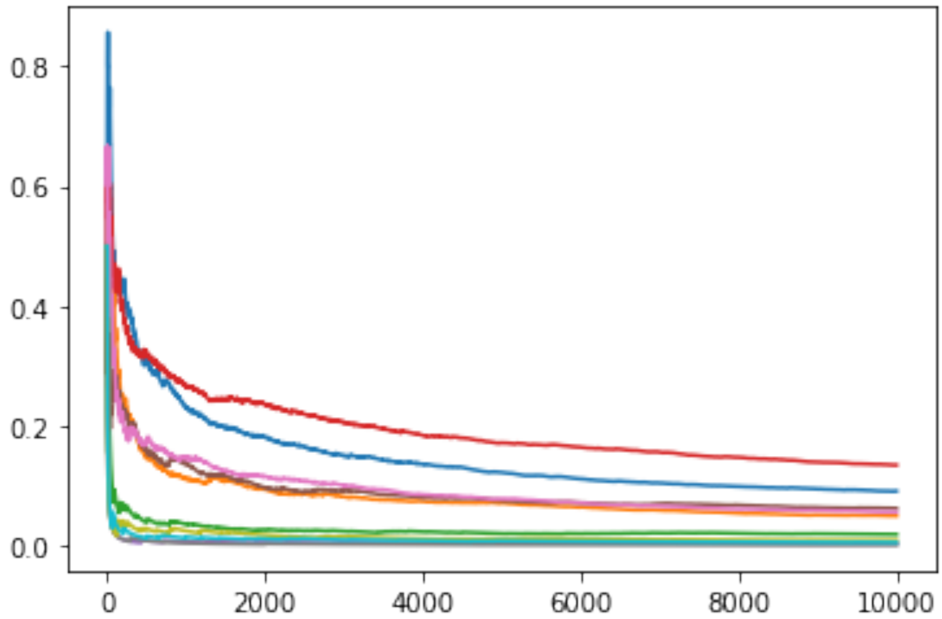
[255]:
```
theta1=0.1
theta2=0.1
q=0.7
n=1000 # number of sample path
t=10000 # number of time steps
x=np.zeros(n)
for i in range(0, n):
    f11=0
    f2=0
    for j in range(0, t):
        p1=(theta1+10*f11)/(theta1+theta2+10*f11+5*f2)
        TypeI=flip(q)
        Med=flip(p1)
        f11+=Med*TypeI
        f2+=(1-Med)
    x[i]=p1
```

[256]:
```
plt.hist(x, bins = 100)
plt.title("histogram for q="+ str(q))
plt.text(q/2, 40, "number of time steps t="+str(t))
plt.show()
```



[289]:
```
theta1=10
theta2=10
q=0.3
n=1000 # number of sample path
m=10000 # number of time steps
x=np.zeros(n)
z=np.zeros((n,m))
for i in range(0, n):
    f11=0
    f2=0
    for j in range(0, t):
        p1=(theta1+10*f11)/(theta1+theta2+10*f11+5*f2)
        z[i,j]=p1
        TypeI=flip(q)
        Med=flip(p1)
        f11+=Med*TypeI
        f2+=(1-Med)
    x[i]=p1
```

[290]:
```
for i in range(0,10):
    plt.plot(z[i,:])
```


0.8
0.6
0.4
0.2
0.0
0
2000
4000
6000
8000
10000

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

M3/4PA48 Dynamics of Games

Solutions to Exercises

24^{th} December, 2020

Contents

1 Replicator Dynamics for One Population

1.1 Nash equilibrium of one population

Exercise 1.1:
1) Let K represent the density of the population of Koalas in Kangaroo Island in South Australia. On one hand, whenever koalas live in areas with an abundance of Eucalyptus leaves (areas of type A) they reproduce at a rate x, whereas they die at a rate y in every other part of the island (areas of type B). The Nash equilibrium of this "game" is simply given by always choosing to live in areas of type A.

2) Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

and suppose we want to study the Nash Equilibria for a game encoded by A. We will do this by looking at the best response function for A. Notice that the matrix A is a permutation matrix, therefore instead of working with entries of (Ax) we can focus on the entries of x directly. Firstly, we compute it for vectors $x \in \Delta$ for which one of the components is bigger than the other two.

Consider $x = (x_1, x_2, x_3)$ where $x_1 > x_2$ and $x_1 > x_3$, then

$$\begin{aligned} \mathcal{BR}(x) &= \arg\max_{y\in\Delta} y \cdot Ax \\ &= \arg\max_{y\in\Delta} (y_1, y_2, y_3) \cdot (x_2, x_3, x_1) \\ &= \arg\max_{y\in\Delta} y_1x_2 + y_2x_3 + y_3x_1 = \{e_3\}. \end{aligned}$$

We can proceed similarly for vectors with $x_2 > x_1, x_3$, and $x_3 > x_1, x_2$.

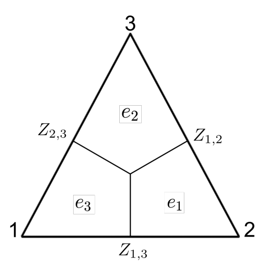


Figure 1: Diagram of $\mathcal{BR}(x)$

The next case we want to consider is the special vector for which all entries are equal, namely $x = (1/3, 1/3, 1/3)$. In this case

$$\mathcal{BR}((1/3, 1/3, 1/3)) = \arg\max_{y\in\Delta} 1/3(y_1 + y_2 + y_3) = \Delta.$$

Finally, we can consider the case where two of the entries of our vectors are equal, and the third one does not dominate (otherwise we go back to the first case we analysed). Essentially we want to study the lines $Z_{ij} = \{x \in \Delta | (Ax)_i = (Ax)_j\} = \{x = (x_1, x_2, x_3) \in \Delta | x_{i+1} = x_{j+1}\}$ for $i, j \in \mathbb{Z}/3\mathbb{Z}$. Suppose we want to find the best response along the segment $Z_{1,2} \cap \{x_1 < 1/3\} = \{x_2 =$

$x_3\} \cap \{x_1 < 1/3\}$, then we would have to maximise $y \cdot Ax$ for $y \in \Delta$, and assuming that $x = (1-a, a/2, a/2)$ and $a \in (2/3, 1)$. Clearly

$$y \cdot Ax = y_1x_2 + y_2x_3 + y_3x_1 = a(y_1 + y_2) + (1-a)y_3,$$

which is maximised whenever $y_1 + y_2 = 1$ and $y_3 = 0$, or $y \in \langle e_1, e_2 \rangle$. Proceeding with a similar reasoning we obtain

$$\begin{aligned}
\mathcal{BR}(Z_{1,2} \cap \{x_1 < 1/3\}) &= \langle e_1, e_2 \rangle \\
\mathcal{BR}(Z_{2,3} \cap \{x_2 < 1/3\}) &= \langle e_2, e_3 \rangle \\
\mathcal{BR}(Z_{1,3} \cap \{x_3 < 1/3\}) &= \langle e_1, e_3 \rangle.
\end{aligned}$$

This means that if we look at the best response over all of Δ we obtain the following

$$\mathcal{BR}(x) = \begin{cases} \{e_3\} & \text{if } x_1 > x_2, x_3 \\ \{e_1\} & \text{if } x_2 > x_1, x_3 \\ \{e_2\} & \text{if } x_3 > x_1, x_2 \\ \langle e_1, e_2 \rangle & \text{if } x \in Z_{1,2} \cap \{x_1 < 1/3\} \\ \langle e_2, e_3 \rangle & \text{if } x \in Z_{2,3} \cap \{x_2 < 1/3\} \\ \langle e_1, e_3 \rangle & \text{if } x \in Z_{1,3} \cap \{x_3 < 1/3\} \\ \Delta & \text{if } x_1 = x_2 = x_3. \end{cases}$$

Now recall that $\hat{x}$ is a Nash Equilibrium if and only if $\hat{x} \in \mathcal{BR}(\hat{x})$. Figure 1 summarises all the information contained in the function above, except for the behaviour on the boundary of every region. It tells us that if there were to Nash Equilbria they would have to belong to either $\partial\Delta$ or $(Z_{1,2} \cap \{x_1 < 1/3\}) \cup (Z_{2,3} \cap \{x_2 < 1/3\}) \cup (Z_{1,3} \cap \{x_3 < 1/3\})$.No vectors in $\langle e_i, e_j \rangle$ are contained in $Z_{i,j} \cap \{x_i < 3\}$, where $i < j \pmod 3$. However, we do have that $(1/3, 1/3, 1/3) \in \Delta = \mathcal{BR}((1/3, 1/3, 1/3))$, hence the vector $(1/3, 1/3, 1/3)$ is the only Nash Equilibrium for A in Δ. It is important to notice that $(1/3, 1/3, 1/3)$ is the meeting point of the three indifference lines $Z_{1,2}$, $Z_{2,3}$, and $Z_{1,3}$.

1.2 Evolutionary stable strategies

Exercise 1.2:
1) We want to determine the Nash Equilibria for

$$A = \begin{pmatrix} 0 & 2 & -1 \\ -1 & 0 & 2 \\ 2 & -1 & 0 \end{pmatrix}$$

It is quite immediate to see that the point $\hat{x} = (1/3, 1/3, 1/3)$ is a Nash Equilibrium:

$$\mathcal{BR}(\hat{x}) = \arg\max_{y \in \Delta} y \cdot A\hat{x} = \arg\max_{y \in \Delta} y \cdot (1/3, 1/3, 1/3) = \Delta \ni \hat{x}.$$

Notice that $A\hat{x} = \hat{x} = (1/3, 1/3, 1/3)$ in accordance with Lemma 1.2 (here $c = 1/3$). We now want to show $\hat{x}$ is an Evolutionary Stable Strategy, and in order to do that we are going to use the second part of Lemma 1.3 in the notes. For any $y \in \Delta$ we have

$$y \cdot Ay = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \cdot \begin{pmatrix} 2y_2 - y_3 \\ -y_1 + 2y_3 \\ 2y_1 - y_2 \end{pmatrix} = y_1y_2 + y_2y_3 + y_1y_3$$
$$\hat{x} \cdot Ay = \frac{2}{3}y_2 - \frac{1}{3}y_3 - \frac{1}{3}y_1 + \frac{2}{3}y_3 + \frac{2}{3}y_1 - \frac{1}{3}y_2 = \frac{1}{3}$$

We now want to show that the function $f(y_1, y_2, y_3) = y_1y_2 + y_2y_3 + y_1y_3$ is maximised at $(1/3, 1/3, 1/3)$ in Δ. In order to do so we will work with Lagrange multipliers. The only constraint we have is that we want to maximum to be in Δ, so the constraint function we will work with is given by $g(y_1, y_2, y_3) = y_1+y_2+y_3-1$. Now we can consider the following $\mathcal{L}(y_1, y_2, y_3, \lambda) = f(y_1, y_2, y_3)-\lambda g(y_1, y_2, y_3) = y_1y_2 + y_2y_3 + y_1y_3 - \lambda y_1 - \lambda y_2 - \lambda y_3 + \lambda$ for $\lambda \in \mathbb{R}$. Therefore, the point of maximum $(\tilde{y_1}, \tilde{y_2}, \tilde{y_3})$ is found by solving the following

$$\nabla\mathcal{L}(\tilde{y_1}, \tilde{y_2}, \tilde{y_3}, \tilde{\lambda}) = \begin{pmatrix} \tilde{y_2} + \tilde{y_3} - \tilde{\lambda} \\ \tilde{y_1} + \tilde{y_3} - \tilde{\lambda} \\ \tilde{y_1} + \tilde{y_2} - \tilde{\lambda} \\ -\tilde{y_1} - \tilde{y_2} - \tilde{y_3} + 1 \end{pmatrix} = 0 \Longrightarrow \begin{pmatrix} \tilde{y_1} \\ \tilde{y_2} \\ \tilde{y_3} \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \\ 2/3 \end{pmatrix}$$

Therefore we know that the maximum of $f(y_1, y_2, y_3)$ in Δ is achieved at $\hat{x}$, and it is precisely $1/3$. So we can conclude that $\hat{x}$ is and ESS, since $y \cdot Ay < \hat{x} \cdot Ay$ for all $y \in \Delta \setminus \{\hat{x}\}$. Since $\hat{x} \in \operatorname{int} \Delta$ is an ESS we can conclude there are no other Nash Equilibria.

Finally $\hat{x}$ is not a strict Nash Equilibrium since for any $y \in \Delta \setminus \{\hat{x}\}$

$$\frac{1}{3} = y \cdot A\hat{x} = \hat{x} \cdot A\hat{x} = \frac{1}{3.}$$

If there was a strict Nash Equilibrium then such a point would be automatically a Nash Equilibrium, contradicting Lemma 1.2.

2) We want to show that e_1, e_2 and e_3 are ESS for

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Define the regions

$$\Xi_i \coloneqq \{x \in \Delta \,|\, x_i > x_{i+1} \text{ and } x_i > x_{i+2}\}$$

for $i \in \mathbb{Z}/3\mathbb{Z}$. Notice that these regions are the same as the ones over which the Best Response function is single-valued (look at the particular shape of A. . .).

Since Ξ_i represents a set of points close to e_i, in order to prove that e_i is an ESS we will show that for all $y \in \Xi_i \setminus \{e_i\}$ the following holds

$$y \cdot Ay < e_i \cdot Ay.$$

Fix $i \in \mathbb{Z}/3\mathbb{Z}$. There are two main observation to make here: if $y \in \Xi_i$ then $y_i > 0$ and $\frac{y_{i+1}}{y_i} < 1$ and $\frac{y_{i+2}}{y_i} < 1$. With this in mind we now have for all $y \in \Xi_i \setminus \{e_i\}$

$$\begin{aligned} y \cdot Ay = |y|^2 &= y_1^2 + y_2^2 + y_3^2 \\ &= y_i(y_i + \frac{y_{i+1}}{y_i} y_{i+1} + \frac{y_{i+2}}{y_i} y_{i+2}) \\ &< y_i(y_i + y_{i+1} + y_{i+2}) \\ &= y_i = e_i \cdot Ay, \end{aligned}$$

as we wanted. We can conclude that e_1, e_2, and e_3 are all ESS.

1.3 Replicator dynamics

Exercise 1.3:
1) We want to study the replicator dynamics described by the matrix

$$A = \begin{pmatrix} 0 & 10 & 1 \\ 10 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

To start with, let us compute the lines

$$Z_{i,j} = \{x \in \Delta | (Ax)_i = (Ax)_j\}$$

for $i, j \in \{1, 2, 3\}$:

$$\begin{aligned} Z_{1,2} &= \{10x_2 + x_3 = 10x_1 + x_3\} \\ &= \{x_1 = x_2\} \\ Z_{2,3} &= \{9x_1 = x_2\} \\ Z_{1,3} &= \{9x_2 = x_1\}. \end{aligned}$$

See Figure 2 for a representation of such indifference lines in Δ. In order to establish the ESS for this system, we will firstly understand its Nash Equilbria, given that every ESS is a NE.

Recall that the intersection of all the indifference lines is a Nash Equilibrium. Given Figure 2 we can see that e_3 is a NE, and that there are no other equilibria in the interior of Δ (since these lines intersect only once).

Analysing the boundary is slightly more delicate. Computing the best response at the corners of the simplex would immediately tell us if such corners are NE or not. The shaded regions in Figure 2 tell us that the best response near e_1 is given by e_2, and that the best response near e_2 is e_1, therefore implying that neither e_1, or e_2 are NE.

Recall that Nash Equilibria along sides are given by intersection with indifference lines. If we consider a side $\langle e_i, e_j \rangle$ then we only need to consider the correspondent indifference line $Z_{i,j}$, and analyse the best response at the intersection point.

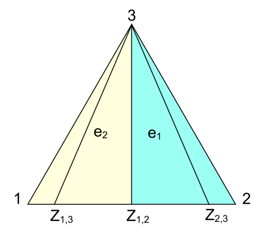


Figure 2: Indifference lines and best response for A.

In our case all the indifference lines intersect the side $\langle e_1, e_2 \rangle$, so we automatically know we will not find Nash Equilibria on $\langle e_2, e_3 \rangle$ and $\langle e_1, e_3 \rangle$, and that we need to focus on the point $(\frac{1}{2}, \frac{1}{2}, 0) = Z_{1,2} \cap \langle e_1, e_2 \rangle$. Such a point is a Nash Equilibrium given that the best response along $Z_{1,2}$ is given by $\langle e_1, e_2 \rangle$ (except at e_3 where it is given by Δ).

The only two NE are given by e_3 and $(\frac{1}{2}, \frac{1}{2}, 0)$. Is e_3 an ESS? In order for e_3 to be an ESS we need that for all $x \in \Delta \setminus \{e_3\}$ and for $\varepsilon > 0$ small enough

$$x \cdot A(\varepsilon x + (1-\varepsilon)e_3) < e_3 \cdot A(\varepsilon x + (1-\varepsilon)e_3).$$

Notice that $x \cdot Ae_3 = 1$ for all $x \in \Delta$, hence the above claim reduces to showing

$$x \cdot Ax < e_3 \cdot Ax$$

for all $x \in \Delta \setminus \{e_3\}$. Now

$$\begin{aligned} x \cdot Ax &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 10x_2 + x_3 \\ 10x_1 + x_3 \\ 1 \end{pmatrix} = 20x_1x_2 + (1 + x_1 + x_2)x_3 \\ e_3 \cdot Ax &= 1 \end{aligned}$$

Consider the vector $\tilde{x} = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} \in \Delta \setminus \{e_3\}$, but then this gives us that $\tilde{x} \cdot A\tilde{x} = 5 > 1 = e_3 \cdot A\tilde{x}$. Therefore, e_3 is NOT an ESS.

Is $\tilde{x} = (\frac{1}{2}, \frac{1}{2}, 0)$ an ESS? For $\tilde{x}$ to be an ESS point we need that

$$y \cdot Ay < \tilde{x} \cdot Ay \tag{1}$$

holds for $y \neq \tilde{x}$ sufficiently close to $\tilde{x}$. Let $y = (\frac{1}{2} + \delta_1, \frac{1}{2} + \delta_2, -\delta_1 - \delta_2)$ where we need to remember that $\delta_1 + \delta_2 \leq 0$, and where δ_1 and δ_2 are assumed not to be zero simultaneously. Then

$$\begin{aligned} y \cdot Ay &= (\frac{1}{2} + \delta_1)(5 + 10\delta_2 - \delta_1 - \delta_2) + (\frac{1}{2} + \delta_2)(5 + 10\delta_1 - \delta_1 - \delta_2) - \delta_1 - \delta_2. \\ &= (\frac{1}{2} + \delta_1)(5 + 9\delta_2 - \delta_1) + (\frac{1}{2} + \delta_2)(5 + 9\delta_1 - \delta_2) - \delta_1 - \delta_2 \\ &= 5 + 8\delta_1 + 8\delta_2 + 18\delta_1\delta_2 - \delta_1^2 - \delta_2^2 \end{aligned}$$

Similarly

$$\tilde{x} \cdot Ay = 5 + 5\delta_1 + 5\delta_2 - \delta_1 - \delta_2 = 5 + 4(\delta_1 + \delta_2).$$

So (1) is equivalent to

$$4(\delta_1 + \delta_2) + 18\delta_1\delta_2 - \delta_1^2 - \delta_2^2 < 0. \tag{2}$$

Without loss of generality, we may assume that $|\delta_2| \le |\delta_1|$ and $\delta_2 = \lambda\delta_1$ with $|\lambda| \le 1$. Now of course we need to remember that $\delta_1 + \delta_2 \le 0$ (and $(\delta_1, \delta_2) \ne (0,0)$) so $(\lambda+1)\delta_1 \le 0$ and therefore either $\delta_1 < 0$ and $\lambda \in (-1, 1]$ or $\delta_1 > 0$ and $\lambda = -1$. So (2) becomes

$$4(\lambda+1)\delta_1 + 18\lambda\delta_1^2 - (\lambda^2+1)\delta_1^2 < 0 \tag{3}$$

with $\delta_1 < 0$ and $\lambda \in (-1, 1]$, or $\delta_1 > 0$ and $\lambda = -1$. If $\delta_1 < 0$ and $\lambda \in (-1, 1]$ then (3) is equivalent to

$$4(\lambda+1) + 18\lambda\delta_1 - (\lambda^2+1)\delta_1 > 0$$

which obviously holds for $|\delta_1|$ small. If $\delta_1 > 0$ and $\lambda = -1$ then (3) is equivalent to

$$-18\delta_1 - 2\delta_1 < 0$$

which again holds. It follows that $\tilde{x}$ is an ESS.

The last thing left to check is the presence of flow singularities. Recall that the replicator dynamics equation is

$$\dot{x_i} = x_i((Ax)_i - x \cdot Ax)$$

for $i \in \{1, 2, 3\}$, which implies

$$\left(\frac{x_i}{x_j}\right)' = \frac{x_i}{x_j}((Ax)_i - (Ax)_j)$$

for $i, j \in \{1, 2, 3\}$. Using the second formulation and for $x \in \Delta$ we have

$$\begin{aligned} \left(\frac{x_1}{x_2}\right)' &= \frac{x_1}{x_2}(10x_2 + x_3 - 10x_1 - x_3) = 10\frac{x_1}{x_2}(x_2 - x_1) \\ \left(\frac{x_3}{x_1}\right)' &= \frac{x_3}{x_1}(x_1 - 9x_2) \\ \left(\frac{x_3}{x_2}\right)' &= \frac{x_3}{x_2}(x_2 - 9x_1). \end{aligned} \tag{4}$$

If we were to have singularities in the interior of Δ then there would exist $x \in \mathrm{int}\Delta$ (notice all its components are in $(0, 1)$) for which

$$\begin{cases} \left(\frac{x_3}{x_1}\right)' = 0 \\ \left(\frac{x_3}{x_2}\right)' = 0 \end{cases} \implies \begin{cases} x_1 - 9x_2 = 0 \\ x_2 - 9x_1 = 0 \end{cases}.$$

Clearly this will never happen for $x_1, x_2 > 0$. Hence the singularities, if they exist, are on $\partial\Delta$. The corners of Δ are all singularity points given the structure of the replicator dynamics ODE: at a corner e_i we have that $\dot{x}_j = 0$ if $j \neq i$ since $x_j = 0$, and $\dot{x}_i = 0$ since $(Ae_i)_i = e_i \cdot Ae_i$. If we look at Equation 4, the only other point at which all equations are zero simultaneously is $\tilde{x} = (\frac{1}{2}, \frac{1}{2}, 0)$, which is the only other singularity on $\partial\Delta$.

2) We now want to investigate how the replicator dynamics for a matrix A changes if we add to its first column the vector $\left(\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right)$. Let

$$B = A + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

and consider the replicator dynamics associated with B

$$\begin{aligned}
\dot{x}_i &= x_i((Bx)_i) - x \cdot Bx) \\
&= x_i \left(\left(Ax + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} x \right)_i - x \cdot Ax - x \cdot \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} x \right) \\
&= x_i \left(\left(Ax + \begin{pmatrix} x_1 \\ x_1 \\ x_1 \end{pmatrix} \right)_i - x \cdot Ax - x_1(x_1 + x_2 + x_3) \right) \\
&= x_i((Ax_i) + x_1 - x \cdot Ax - x_1) \\
&= x_i((Ax)_i - x \cdot Ax).
\end{aligned}$$

The replicator dynamics of B is the same as the replicator dynamics of A.

3) The RHS of the replicator dynamics equation is $\mathcal{C}^\infty$-regular in every entry, therefore we can apply the local Picard–Lindelöf Theorem and obtain that for any starting point $x(0) \in \Delta$ we have a unique local solution. Notice that the solution can never leave the simplex, hence its norm is always bounded. This implies that the solution cannot blow up in finite time, therefore the unique local solution we have precedently established exists for all times.

1.4 ESS points are asymptotically stable for the replicator system

Exercise 1.4:
Consider the matrix $A = \mathbf{Id}_3$, the 3×3 identity matrix. For $\hat{x} = e_j$ and $j \in \{1, 2, 3\}$, consider the function

$$P(x) = \prod_{i=1}^{3} x_i^{\hat{x}_i} = \prod_{i=1}^{3} x_i^{(e_j)_i} = x_j$$

where $x \in \Delta$. We will show that the flow tends to e_j for each $j \in \{1,2,3\}$. In order to simplify the calculations let $\hat{x} = e_1$, so that $P(x) = x_1$. Then

$$\begin{aligned}\frac{\dot{P}}{P}(x) &= \hat{x} \cdot Ax - x \cdot x \\ &= e_1 \cdot \mathbf{Id}_3 x - x \cdot \mathbf{Id}_3 x \\ &= e_1 \cdot x - |x|^2 = x_1 - x_1^2 - x_2^2 - x_3^2\end{aligned}$$

If now let x be near e_1, then we can write it as $x = \left(\begin{smallmatrix} 1-\varepsilon \\ \delta \\ \tau \end{smallmatrix}\right)$ for $\varepsilon, \delta, \tau \in (0,1)$, and $\delta + \tau = \varepsilon$ so that

$$\begin{aligned}\frac{\dot{P}}{P}(x) &= 1 - \varepsilon - (1-\varepsilon)^2 - \delta^2 - \tau^2 \\ &= \varepsilon - \varepsilon^2 - \delta^2 - \tau^2 \\ &> \varepsilon - \varepsilon^2 - \varepsilon^2 \qquad \text{since } \delta^2 + \tau^2 < \epsilon^2 \\ &= \varepsilon(1 - 2\varepsilon) > 0 \quad \text{since } \varepsilon > 0 \text{ (and small).}\end{aligned}$$

Therefore, $\dot{P}(x) > 0$ for $x \in \Delta \setminus \{e_1\}$ close to e_1, so e_1 attracts nearby points. Similar computations, where we take $\hat{x} = e_2$ or e_3 in the definition of P, show that the vertices of Δ attract nearby points.

Next, we turn our attention to the boundary of Δ. So thanks to the replicator dynamics equation

$$\left(\frac{x_i}{x_j}\right)' = \frac{x_i}{x_j}(x_i - x_j)$$

where $i, j \in \{1,2,3\}$. For example along $\langle e_1, e_2 \rangle$ we have that the sign of $\left(\frac{x_1}{x_2}\right)'$ changes at $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$. Similarly, we have a sign change at $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$ and $\left(0, \frac{1}{2}, \frac{1}{2}\right)$.

These same equations tell us that along the indifference line $Z_{i,j}$ the derivative $\left(\frac{x_i}{x_j}\right)'$ is equal to 0, and the direction of the flow is completely described by the derivative in the third component. Let us make an example. Suppose we want to consider the flow along $Z_{1,2}$, then we know that $\left(\frac{x_1}{x_2}\right)' = 0$, so we consider $\dot{x}_3$. Any $x \in Z_{1,2}$ can be written as $x = \left(\begin{smallmatrix} \frac{1-x_3}{2} \\ \frac{1-x_3}{2} \\ x_3 \end{smallmatrix}\right)$, and the replicator equation gives us

$$\begin{aligned}\dot{x}_3 &= x_3((\mathbf{Id}_3 x)_3 - x \cdot \mathbf{Id}_3 x) \\ &= x_3(x_3 - |x|^2) \\ &= x_3(x_3 - 2\left(\frac{1-x_3}{2}\right)^2 - x_3^2) \\ &= -\frac{1}{2}x_3(3x_3^2 - 4x_3 + 1) = -\frac{1}{2}x_3(x_3 - 1)(x_3 - \frac{1}{3}).\end{aligned}$$

We can conclude that $\dot{x}_3 < 0$ whenever $x_3 \in \left(0, \frac{1}{3}\right)$, and that $\dot{x}_3 > 0$ for $x_3 \in \left(\frac{1}{3}, 1\right)$. The same computations show identical behaviour along $Z_{1,3}$ and $Z_{2,3}$.

Consider now the Nash Equilibrium $\hat{x} = (1/3, 1/3, 1/3)$. How does the flow behave around it? From the previous analysis we have carried out for the flow along $Z_{i,j}$ we expect the flow to be repelled by $\hat{x}$ (star node). In order to see if our hunch is correct we will linearise the RHS of the replicator equation at $\hat{x}$. There are different equivalent ways to do so, for example see Example 1.8 in the notes. We will take a more direct approach.

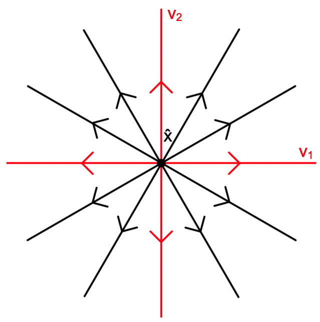


Figure 3: Linearisation around $\hat{x}$.

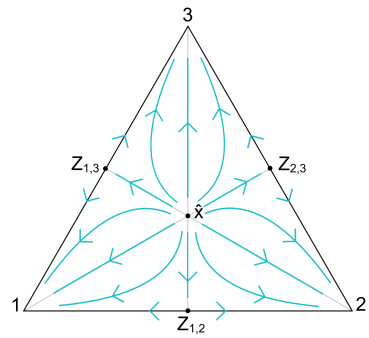


Figure 4: Flow in Δ for $A = \mathbf{Id}_3$.

Let $h \in \mathbb{R}^3$ be a vector whose entries sum up to 0, i.e. $\sum_{i=1}^{3} h_i = 0$, and consider the perturbed vector $p = \hat{x} + h$. Now let $i, j, k \in 1, 2, 3$ be distinct

$$\begin{aligned}\dot{h_i} = \dot{p}_i &= p_i((\mathbf{Id}_3 p)_i - p \cdot \mathbf{Id}_3 p) \\ &= (\frac{1}{3} + h_i)(\frac{1}{3} + h_i - \sum_{l=1}^{3}(\frac{1}{3} + h_l)^2) \\ &= (1/3 + h_i)(\frac{1}{3}h_i - \frac{2}{3}h_j - \frac{2}{3}h_k + \mathcal{O}(h^2)) \\ &= (\frac{1}{3} + h_i)(h_i + \mathcal{O}(h^2)) \qquad \text{since } h_i = -h_j - h_k \\ &= \frac{1}{3}h_i + \mathcal{O}(h^2).\end{aligned}$$

Therefore, the linearisation yields the matrix

$$L = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$$

with repeated eigenvalues $\lambda_1 = \lambda_2 = \frac{1}{3}$ and associated eigenvectors $v_1 = \binom{1}{0}$, and $v_2 = \binom{0}{1}$. As we suspected, the point $\hat{x}$ is a star node. See Figure 3 for a representation of the flow near $\hat{x}$ and Figure 4 for the flow in Δ.

1.5 Further examples

Exercise 1.5:
We want to establish the phase portrait for the replicator equation where

$$A = \begin{pmatrix} 0 & 10 & 1 \\ 10 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

We already know, thanks to Exercise 1.3 that this system has two Nash Equilibria, namely e_3, and $\tilde{x} = (\frac{1}{2}, \frac{1}{2}, 0)$, and flow singularities at each vertex of Δ and at $\tilde{x}$.

Let us study what happens along the boundary of Δ. Let $x \in \langle e_1, e_3 \rangle$, so that $x = \left(\begin{smallmatrix} x_1 \\ 0 \\ x_3 \end{smallmatrix}\right)$ where $x_1 = 1 - x_3$ for $x_3 \in (0,1)$. Along such a side we have

$$\begin{aligned} \left(\frac{x_3}{x_1}\right)' &= \left(\frac{x_3}{x_1}\right)(x_1 - 9x_2) \\ &= \frac{x_3}{x_1} x_1 = x_3 > 0 \end{aligned}$$

which means that the flows goes from e_1 towards e_3. Along $\langle e_2, e_3 \rangle$ we see that $\left(\frac{x_3}{x_2}\right)' > 0$, therefore the solution flows from e_2 towards e_3.

The last side contains a singularity, hence we expect a more interesting behaviour. For $x = \left(\begin{smallmatrix} x_1 \\ x_2 \\ 0 \end{smallmatrix}\right) \in \langle e_1, e_2 \rangle$, where $x_1 = 1 - x_2$ for $x_2 \in (0,1)$, we obtain

$$\left(\frac{x_1}{x_2}\right)' = 10\frac{x_1}{x_2}(x_2 - x_1) = 10\frac{(1-x_2)(2x_2-1)}{x_2}.$$

Such a function is negative between 0 and $\frac{1}{2}$, zero at $\frac{1}{2}$ (as we expected by flow singularities), and positive between $\frac{1}{2}$ and 1, so the flows is attracted by $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$ from e_1 and e_2. See Figure 5 for more details.

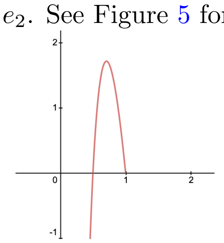


Figure 5: Graph of $(\frac{x_1}{x_2})'$ along $\langle e_1, e_2 \rangle$.

Along $Z_{2,3}$	Along $Z_{1,2}$	Along $Z_{1,3}$
$\left(\frac{x_3}{x_1}\right)' < 0$	$\left(\frac{x_3}{x_1}\right)' < 0$	$\left(\frac{x_3}{x_1}\right)' = 0$
$\left(\frac{x_3}{x_2}\right)' = 0$	$\left(\frac{x_3}{x_2}\right)' < 0$	$\left(\frac{x_3}{x_2}\right)' < 0$
$\left(\frac{x_1}{x_2}\right)' > 0$	$\left(\frac{x_1}{x_2}\right)' = 0$	$\left(\frac{x_1}{x_2}\right)' < 0$

Table 1: Flow along indifference lines.

A similar analysis can be carried out along the indifference lines $Z_{1,2}, Z_{2,3}, Z_{1,3}$ as summarised in Table 1. Therefore, the flow along the indifference lines leaves e_3 and goes towards $\langle e_1, e_2 \rangle$.

As we have proved in Exercise 1.3 part 1, the point $\tilde{x}$ is an ESS, therefore it is asymptotically stable (as the flow analysis we have just carried would suggest).

1.6 Rock–Paper–Scissors replicator game

Exercise 1.6:
1) We want to model the game of Rock–Paper–Scissors. In such game we have three strategies R, P, S. The rules are quite easy R beats S, which beats P, which beats R, and any strategy played against itself resolves to a draw. See Figure 6.

	Rock	Paper	Scissors
Rock	0	+1	-b
Paper	-b	0	1
Scissors	1	-b	0

Table 2: Payoff table.

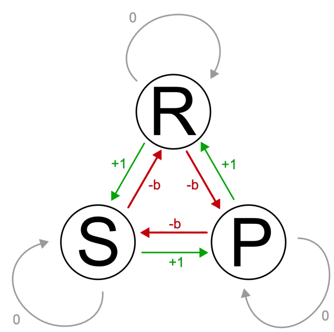


Figure 6: A schematics of the interactions between the various strategies.

The payoff for winning is $+1$, for losing $-b$, where $b > 0$, and for drawing 0. If we write all such data into a table we obtain Table 2. From this, we can read off the payoff matrix

$$A = \begin{pmatrix} 0 & +1 & -b \\ -b & 0 & 1 \\ 1 & -b & 0 \end{pmatrix}.$$

2) Consider the three vectors

$$\begin{aligned} A_1 &= \frac{1}{1+b+b^2}(1, b^2, b)^\mathsf{T} \\ A_2 &= \frac{1}{1+b+b^2}(b, 1, b^2)^\mathsf{T} \\ A_3 &= \frac{1}{1+b+b^2}(b^2, b, 1)^\mathsf{T} \end{aligned}$$

from Lemma 1.5. We claim that A_i, A_{i+1}, and e_{i+1} are collinear (where the indexes are to be taken in $\mathbb{Z}/\mathbb{Z}_3$). We will only show the computations for A_3, A_1, e_1, but every other case is identical. Recall that three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are collinear if and only if $\mathbf{a}-\mathbf{c}$, and $\mathbf{b}-\mathbf{c}$ are parallel. Therefore

$$A_3 - e_1 = \frac{1}{1+b+b^2}\begin{pmatrix}-1-b\\ b\\ 1\end{pmatrix}$$
$$A_1 - e_1 = \frac{b}{1+b+b^2}\begin{pmatrix}-1-b\\ b\\ 1\end{pmatrix}$$

which means that $A_1 - e_1 = b(A_3 - e_1)$, so A_3, A_1, e_1 are collinear.

3) At the beginning of the proof of Lemma 1.5 we say that

$$\frac{1}{T}\int_0^T x(t)\cdot Ax(t)dt \to 0 \qquad \text{as } T \to +\infty, \tag{5}$$

let us show why this is true. Recall that

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

where $b > 1$. The reasons why this happens are sketched in the proof, and they are basically two

1. The payoff $x \cdot Ax$ tends to 0 as the flows $x(t)$ gets closer and closer to any vertex of Δ;
2. $x(t)$ spends most of the time close to the vertices of Δ.

To deal with 1) consider a point

$$x = \begin{pmatrix}1-\varepsilon\\ \delta\\ \tau\end{pmatrix}$$

in Δ close to e_1, where $0 < \delta, \tau \leq \epsilon < 1$. Now

$$\begin{aligned}|x \cdot Ax| &= \left|\begin{pmatrix}1-\varepsilon\\ \delta\\ \tau\end{pmatrix}\cdot\begin{pmatrix}0 & 1 & -b\\ -b & 0 & 1\\ 1 & -b & 0\end{pmatrix}.\begin{pmatrix}1-\varepsilon\\ \delta\\ \tau\end{pmatrix}\right| =\\
&= |-b\varepsilon(1-\varepsilon) + (1-b)\delta(1-\delta)|\\
&\leq b\varepsilon(1-\varepsilon) + (1-b)\delta(1-\delta)\\
&< b\varepsilon + (b-1)\varepsilon\\
&= \varepsilon(2b-1)\end{aligned}$$

which means that $x \cdot Ax \to 0$ as $\varepsilon \to 0$. So the closer we are to the vertices of Δ, the closer to zero the payoff is.

Next, we want to show that the speed of the flow is almost zero near the vertices of Δ, and maximal away from them. Consider the side of Δ between e_1 and e_2, and let $x = x_1 e_1 + (1 - x_1)e_2$ be a point on it, where $x_1 \in [0, 1]$. Now

$$Ax = \begin{pmatrix} 1 - x_1 \\ -bx_1 \\ x_1 - b(1 - x_1) \end{pmatrix}$$
$$x \cdot Ax = x_1(1 - x_1)(1 - b)$$

so we have, thanks to the replicator equation

$$\begin{aligned} |\dot{x}|^2 &= x_1^2((1 - x_1) - x_1(1 - x_1)(1 - b))^2 + (1 - x_1)^2(-bx_1 - x_1(1 - x_1)(1 - b))^2 \\ &= x_1^2(1 - x_1)^2[(1 - x_1(1 - b))^2 + (b + (1 - x_1)(1 - b))^2]. \end{aligned}$$

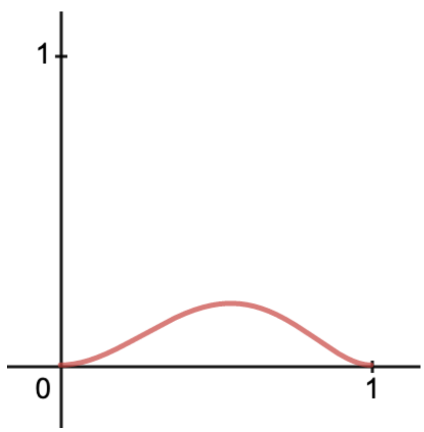


Figure 7: Graph of $|\dot{x}|^2$ for $b = 1.5$.

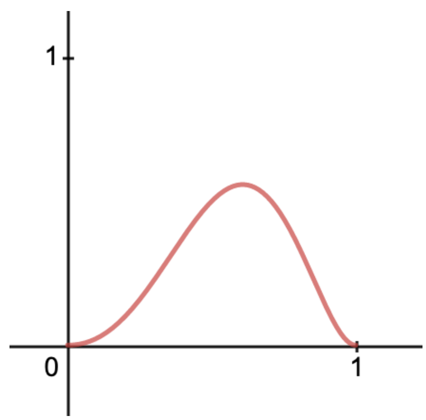


Figure 8: Graph of $|\dot{x}|^2$ for $b = 3$.

Thanks to continuity of the solution of ODEs, we can expect the velocity of the flow in int Δ close to the boundary to behave like $|\dot{x}|$ (as computed above). This function has a (local) maximum in (0,1) and tends to zero as x_1 approaches 0 or 1, i.e. the flow has maximal velocity away from the corners of Δ and gets smaller as it gets closer to them (see Figure 7 and 8 to get an idea of what this function looks like along on of the sides of Δ).

We have just established that the payoff of this game gets small near the corners of Δ, and that in these areas the flow has low speed, meaning that it spends most of the time there. Intuitively this is why the limit 5 converges to 0, but in order to give a rigorous proof of this we have to estimate how much time we spend in the corners (at least for T big enough).

We will now assume that Δ is our ambient space, and we equip it with the subset topology coming from the Euclidean topology of $\mathbb{R}^3$, i.e. all the neighbourhoods we will consider from now on are neighbourhoods in Δ. Consider the neighbourhood

$$\Omega = \bigcup_{i=1}^{3} \Omega_i = \bigcup_{i=1}^{3} B_\varepsilon(e_i)$$

of the vertices of Δ, where $B_\varepsilon(e_i)$ is the ball of radius ε around e_i. Similarly, consider the ε^2-tubular neighbourhood of $\partial\Delta$ given by

$$\Xi = \overline{B_{\varepsilon^2}(\partial\Delta)} \setminus \Omega = \overline{\bigcup_{x\in\partial\Delta} B_{\varepsilon^2}(x)} \setminus \Omega.$$

Notice that $\varepsilon > 0$ can be taken small enough so that we can apply the Hartman–Grobman theorem to $B_\varepsilon(e_i)$ for $i = 1, 2, 3$, and such that the payoff $|x \cdot Ax|$ is bounded above by ε over Ω.

As we wrote before Ω is made up by 3 components $\Omega_i = B_\varepsilon(e_i)$, and similarly Ξ is made up by 3 components. We will call Ξ_i the components in which the flows travels from Ω_{i-1} to Ω_i, where all the indexes have to be take $\mod 3$. Notice that Ξ is compact, and the flow over this set has always non-zero derivative since it is away from e_1, e_2, e_3. We can define $C \coloneqq \min_\Xi |\dot{x}|$ which tells us that $\max T_{\Xi_i} = \frac{1-2\varepsilon}{C} = K$, for T_{Ξ_i} being the time it takes the flow to get through Ξ_i. In order to estimate $\max T_{\Xi_i}$ we have used small angle approximations, linearised estimates of the flow, and we maximised the equation Time=Displacement/Speed. As time $T \to \infty$ we can see that $\max T_{\Xi_i}$ remains constant ($= K$), meaning that the maximal time for the flow to get through Ξ_i is constant, and does not depend on how close the flow gets to $\partial\Delta$.

If we denote by T^N the amount of time that the flow takes to complete a full loop then we can break this down as

$$T^N = T^N_{\Omega_1} + T^N_{\Xi_2} + T^N_{\Omega_2} + T^N_{\Xi_3} + T^N_{\Omega_3} + T^N_{\Xi_1} \leq T^N_{\Omega_1} + T^N_{\Omega_2} + T^N_{\Omega_3} + 3K = T^N_{\Omega} + 3K$$

where $T^N_{\Omega_i}$, and $T^N_{\Xi_i}$, represent the time needed to get through Ω_i, Ξ_i respectively during the loop N. We now will proceed to show that

$$\lim_{N\to\infty} \frac{T^{N+1}}{T^N} = \lim_{N\to\infty} \frac{T^{N+1}_\Omega}{T^N_\Omega} < \infty.$$

In order to compute such a limit we want estimate the time it takes the flow to traverse Ω_i. Let us start by considering Ω_1. By applying the Hartman-Grobman theorem to Ω_1 we have that the flow generated by the replicator equation is $\mathcal{C}^1$ conjugated to the linearised flow given by

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} 1 & 1+b \\ 0 & -b \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}. \tag{6}$$

Notice that the eigenvalues of the matrix are $\lambda_1 = 1$ and $\lambda_2 = -b$ with associated eigenvectors $v_1 = \binom{1}{0}$, and $v_2 = \binom{1}{-1}$, respectively.

Remark. *The Hartman-Grobman theorem states that an equilibrium of a system of ODEs is locally linearisable as long as the linearisation matrix L is hyperbolic (all its eigenvalues have non-zero real part). If the flow is 2 dimensional then the conjugacy between the flow given by the original system and the one given by $\dot{\mathbf{x}} = L\mathbf{x}$ is $\mathcal{C}^1$ regular. For more general flows then the conjugacy is only α-Hölder continuous, where α depends on the eigenvalues of L.*

Let X, Y be two subsets of Euclidean spaces, and $r \in \mathbb{N}$. We say two flows $\varphi : X \to X$, and $\psi : Y \to Y$ are $\mathcal{C}^r$ conjugated if there exists a $\mathcal{C}^r$ diffeomorphism $h : X \to Y$ such that $h \circ \varphi = \psi \circ h$. If the last equality only holds over a subset of X then we say that φ and ψ are locally $\mathcal{C}^r$ conjugated.

By making a simple change of basis transformation we can (smoothly) conjugate the flow given by Equation 6 to

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} -b & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}. \tag{7}$$

It is immediate to see that the eigenvalues of the matrix are given by $\lambda_1 = -b$, and $\lambda_2 = 1$, but now the associated eigenvectors are $v_1 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, and $v_2 = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ respectively. The solution to this system of ODEs for $\left(\begin{smallmatrix} x(0) \\ y(0) \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right)$ is given by

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x_0 e^{-bt} \\ y_0 e^t \end{pmatrix}.$$

As we have anticipated we want to understand how long it takes our original flow to cross Ω_1, and while doing so we will see how quickly the flow tends to $\partial\Delta$ in terms of distance from it. Let $\left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} \rho \\ \tau \end{smallmatrix}\right)$ for $\rho, \tau > 0$ arbitrarily small. We want to establish how long it takes to get to $\left(\begin{smallmatrix} \tau' \\ \rho \end{smallmatrix}\right)$, and what is the size of τ' compared to τ. This is done by solving

$$\begin{pmatrix} \tau' \\ \rho \end{pmatrix} = \begin{pmatrix} \rho e^{-bt} \\ \tau e^t \end{pmatrix}$$

which gives us $t = \ln \frac{\rho}{\tau}$ and $\tau' = \rho^{1-b}\tau^b = C_1\tau^b$. If we fix ρ and we let $t \to \infty$ we get $\tau \to 0$ which tells us that the flow tends to $\partial\Delta$, as we know. This estimates have been computed for System 7, which is $\mathcal{C}^1$ conjugated to the replicator dynamics, so thanks to the regularity of the conjugacy we know that the same asymptotics hold for the original flow. This means that for N big enough, and if $d \leq \varepsilon^2$ denotes the distance between the flow entering Ω_1 to start the N^{th} loop then the time to exit Ω_1 is $\sim \ln \frac{\varepsilon}{d}$, and that time the flow will be $\sim d^b$ away from $\partial\Delta$. These asymptotic estimates only depend on the eigenvalues of the linearisation of the flow at e_1, by proceeding similarly one can show that the situation in Ω_2, and Ω_2 is identical. We will assume that the distance flow-$\partial\Delta$ when entering Ω_{i+1} is approximately the same as when leaving Ω_i. The time to complete the loop is approximately given by

$$\begin{aligned}T^N \sim T^N_\Omega &= T^N_{\Omega_1} + T^N_{\Omega_2} + T^N_{\Omega_3}\\ &\sim \ln\frac{\varepsilon}{d} + \ln\frac{\varepsilon}{d^b} + \ln\frac{\varepsilon}{d^{b^2}}\\ &= \ln\frac{\varepsilon^3}{d^{1+b+b^2}}.\end{aligned}$$

Note that when the flow comes back to Ω_1 its distance to $\partial\Delta$ is $\sim d^{b^3}$, hence $T^{N+1} \sim \ln\frac{\varepsilon}{d^{b^3+b^4+b^5}}$. Since we have expressed T^N in terms of d then taking a limit as $N \to \infty$ is the same as $d \to 0$. Therefore,

$$\begin{aligned}\lim_{N\to\infty}\frac{T^{N+1}}{T^N} &= \lim_{d\to 0}\frac{\ln\frac{\varepsilon^3}{d^{b^3+b^4+b^5}}}{\ln\frac{\varepsilon^3}{d^{1+b+b^2}}}\\ &= \lim_{d\to 0}\frac{\ln\frac{1}{d^{b^3+b^4+b^5}}}{\ln\frac{1}{d^{1+b+b^2}}} = b^3\end{aligned}$$

So for N large enough $T^{N+1} \sim b^3 T^N$, where $b > 1$.

If we assume that T is big enough, we have that T can be approximately written as the sum of the times it takes to do N loops, or equivalently

$$T = \sum_{i=0}^{N-1} T^i \sim T^0 \sum_{i=0}^{N-1} b^{3i} = T^0\frac{1-b^{3N}}{1-b^3}.$$

This approximation allows us to compute how many loops we expect to have completed in a fixed (large) time T

$$N = \frac{\ln\left(\left(\frac{b^3-1}{T_0}\right)T + 1\right)}{3\ln b}$$

which immediately gives us that

$$\lim_{T\to\infty}\frac{N}{T} = 0.$$

We can finally prove the limit in Equation 5. Therefore

$$\begin{aligned}\lim_{T\to\infty}\left|\frac{1}{T}\int_0^T x\cdot Ax dt\right| &\leq \lim_{T\to\infty}\frac{1}{T}\int_0^T |x\cdot Ax| dt\\ &= \lim_{T\to\infty}\frac{1}{T}\left(\int_{\sum_{j=0}^N T^j_\Omega}|x\cdot Ax|dt + \int_{\sum_{j=0}^N T_\Xi}|x\cdot Ax|dt\right)\\ &\leq \lim_{T\to\infty}\frac{\sum_{j=0}^N T^j_\Omega}{T}\varepsilon + \frac{N}{T}\max_\Xi|x\cdot Ax| \leq \varepsilon\end{aligned}$$

since $\frac{\sum_{j=0}^N T^j_\Omega}{T} \sim \frac{\sum_{j=0}^N T^j}{T} \to 1$ and $\frac{N}{T} \to 0$. Since ε can be taken arbitrarily small we are done.

1.7 Hypercycle equation and permanence

Exercise 1.7:
We want to show that $\gamma_k = \sum_{j=0}^{n-1} c_j \lambda^{jk}$, where $k = 0, 1, \ldots, n$ and $\lambda = e^{\frac{2\pi i}{n}}$, are the eigenvalues of

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix}.$$

We know, by the proof of Lemma 1.6, that the corresponding eigenvector to γ_k is

$$v_k = \begin{pmatrix} 1 \\ \lambda^k \\ \vdots \\ \lambda^{(n-1)k} \end{pmatrix}.$$

It is just a matter of multiplying C and v_k, and show that the product equals $\gamma_k v_k$. For $k = 0, 1, \ldots, n-1$ we have

$$\begin{aligned} Cv_k &= \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix} \begin{pmatrix} 1 \\ \lambda^k \\ \vdots \\ \lambda^{(n-1)k} \end{pmatrix} \\ &= \begin{pmatrix} c_0 + c_1\lambda^k + \ldots + c_{n-1}\lambda^{(n-1)k} \\ \lambda_k(c_{n-1}\lambda^{(n-1)k} + c_0 + \ldots + c_{n-2}\lambda^{(n-2)k}) \\ \vdots \\ \lambda^{(n-1)k}(c_1\lambda^k + c_2\lambda^{2k} + \ldots + c_0) \end{pmatrix} \\ &= \sum_{j=0}^{n-1} c_j\lambda^{jk} \begin{pmatrix} 1 \\ \lambda^k \\ \vdots \\ \lambda^{(n-1)k} \end{pmatrix} = \gamma_k v_k. \end{aligned}$$

1.8 Existence and the number of Nash Equilibria

Exercise 1.8
1) We want to check the Poincaré-Hopf formula holds for simple flows X on some surface M. Recall that the formula states

$$\sum_{\substack{x \in M \\ X(x)=0}} i_X(x) = \chi(M)$$

where $i_X(x)$ is the index of X at x and $\chi(M)$ is the Euler characteristic of M.

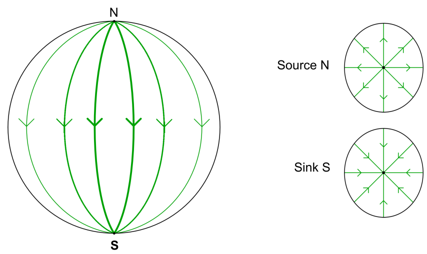


Figure 9: North-South flow on $\mathbb{S}^2$.

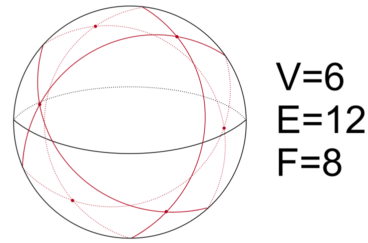


Figure 10: A triangulation of $\mathbb{S}^2$.

Let $M = \mathbb{S}^2$ be the two dimensional sphere, and consider the north-south flow X on it. As we can see from Figure 9 this flow has to singularities at N (the north pole) and S (the south pole). N is a source, whereas S is a sink, which means that $i_X(N) = i_X(S) = +1$.

Recall that the Euler characteristic of a surface M can be computed as

$$\chi(M) = V - E + F$$

where V is the number of vertices, E the number of edges, and F the number of faces of a triangulation of M. The Euler characteristic is independent from the choice of triangulation (as long as you are not collapsing triangles). For more information about Euler characteristic and triangulations look up CW-complexes or simplicial complexes.

The triangulation of $\mathbb{S}^2$ in Figure 10 tells us that $\chi(\mathbb{S}^2) = 6 - 12 + 8 = +2$. Now if we put everything together

$$\sum_{\substack{x \in \mathbb{S}^2 \\ X(x)=0}} i_X(x) = i_X(N) + i_X(S) = 1 + 1 = 2 = \chi(\mathbb{S}^2).$$

Let us consider a different surface. Let M be the two dimensional torus $\mathbb{T}^2$, and let X be the north-south flow as in Figure 11.

The flow X has now 4 singularities, namely a source N, a sink S and two saddle points N', S'. Their indexes are

$$\begin{aligned} i_X(N) &= +1; \\ i_X(S') &= -1; \\ i_X(N') &= -1; \\ i_X(S) &= +1. \end{aligned}$$

The Euler characteristic of $\mathbb{T}^2$ is readily computed thanks to the triangulation shown in Figure 12

$$\chi(\mathbb{T}^2) = 9 - 27 + 18 = 0.$$

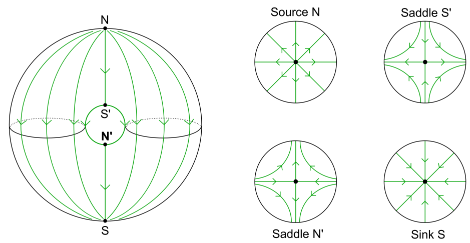


Figure 11: North-South flow on $\mathbb{T}^2$.

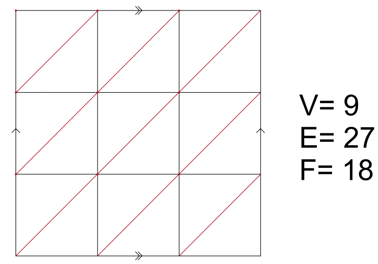


Figure 12: A triangulation of $\mathbb{T}^2$.

Now let us put everything together as before we have

$$\sum_{\substack{x \in \mathbb{T}^2 \\ X(x)=0}} i_X(x) = i_X(N) + i_X(S') + i_X(N') + i_X(S) = 1 - 1 - 1 + 1 = 0 = \chi(\mathbb{T}^2).$$

2) See Example 1.19 in the notes.

3) Consider, for $\varepsilon > 0$, the perturbed flow

$$\dot{x}_i = x_i((Ax)_i - x \cdot Ax) + \varepsilon$$

and assume that the original flow only presents regular singularities (the linearisation matrix at the singularity is invertible). We claim that under this assumption all the Nash Equilibria of the original flow on the boundary move towards the interior of the simplex under the perturbed flow, and the other singularities of the system move outwards.

Let us assume that Δ is a 3 dimensional simplex in order to simplify the discussion. It is important to notice that the singularity of the original replicator equation remain the same under the perturbed flow and they move smoothly as ε varies. The assumption that every singularity is regular implies that there are only finitely many singularities, and that they are all isolated.

Assume that $\hat{x}$ is a Nash Equilibrium on $\partial\Delta$ for the original game. Hence there exists $i \in \{1,2,3\}$ for which $\dot{\hat{x}}_i = 0$. As was shown in the proof of Theorem 1.4, the i-th component of the vector field $X_\varepsilon(x)$ takes the form $x_i z_i + \epsilon + \text{h.o.t.}$ (higher order terms) where $z_i = (A\hat{x})_i - \hat{x} \cdot A\hat{x}$. Since $z_i < 0$ when $\hat{x}$ is a Nash equilibrium, the singularity $\hat{x}_\epsilon$ for $X_\varepsilon(x)$ near $\hat{x}$ has a positive i-th component (and so moves to the interior of Δ). Similarly, if $\hat{x}$ is not a Nash Equilibrium then $z_i \geq 0$, but since we have assumed that the vector field is regular, we have $z_i > 0$. It follows that in this case $\hat{x}$ moves to outside Δ.

2 Two Players Games

2.1 Two conventions for the payoff matrices

Exercise 2.1:
1) Consider the two players game given by the matrix

$$G = \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (-1,1) \end{pmatrix}$$

where we have adopted the 2^{nd} convention (see the notes). From G we can read off the two matrices determining this game

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad B = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

and we can use them to compute the best response for both players. Recall that $(\hat{x}, \hat{y})$ is a Nash Equilibrium if and only if $\hat{x} \in \mathcal{BR}_A(\hat{y})$ and $\hat{y} \in \mathcal{BR}_B(\hat{x})$. For $x, y \in \Delta = \langle e_1, e_2 \rangle$

$$\begin{aligned}
\mathcal{BR}_A(y) &= \underset{x \in \langle e_1, e_2 \rangle}{\arg\max}\, x \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} y \\
&= \underset{x \in \langle e_1, e_2 \rangle}{\arg\max}\, x_1 y_1 - x_2 y_2 = \{e_1\} \\
\mathcal{BR}_B(x) &= \underset{y \in \langle e_1, e_2 \rangle}{\arg\max}\, x \cdot \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} y \\
&= \underset{x \in \langle e_1, e_2 \rangle}{\arg\max}\, -x_1 y_1 + x_2 y_2 = \{e_2\}
\end{aligned}$$

Hence we can immediately see that $e_1 \in \mathcal{BR}_A(e_2)$ and $e_2 \in \mathcal{BR}_B(e_1)$, so (e_1, e_2) is the only Nash Equilibrium for G.

2) Consider a two-person game (A, B), and denote by $\Delta_A \times \Delta_B$ its phase space. Let $(\hat{x}, \hat{y}) \in \text{int } \Delta_A \times \Delta_B$ be a Nash Equilibrium for the game (A, B). If we work with the second notation then we know that $\hat{x}$ maximises the product $x \cdot A\hat{y}$ for $x \in \Delta_A$, and that $\hat{y}$ maximises the product $\hat{x} \cdot By$ for $y \in \Delta_B$. Therefore, for any i, j we have

$$e_i \cdot A\hat{y} \leq \hat{x} \cdot A\hat{y} \qquad \hat{x} \cdot Be_j \leq \hat{x} \cdot B\hat{y}.$$

The two vectors $\hat{x}, \hat{y}$ can be written as the linear combinations $\hat{x} = \sum_i \lambda_i e_i$, and $\hat{y} = \sum_j \rho_j e_j$, where $\lambda_i, \rho_j > 0$ for all i, j since we have assumed that the Nash Equilibrium is contained in the interior of our phase space, and $\sum_i \lambda_i = \sum_j \rho_j = 1$ since we are working with probability vectors. Then if we sum over the two previous inequalities we obtain

$$\begin{aligned}
\hat{x} \cdot A\hat{y} &= \sum_i \lambda_i e_i \cdot A\hat{y} \leq \sum_i \lambda_i \hat{x} \cdot A\hat{y} = \hat{x} \cdot A\hat{y} \\
\hat{x} \cdot B\hat{y} &= \sum_j \rho_j \hat{x} \cdot Be_j \leq \sum_j \rho_j \hat{x} \cdot B\hat{y} = \hat{x} \cdot B\hat{y}.
\end{aligned}$$

In order to get a strict inequality in the previous derivation we would need at least one i and/or j such that $e_i \cdot A\hat{y} < \hat{x} \cdot A\hat{y}$, and/or $\hat{x} \cdot Be_j < \hat{x} \cdot B\hat{y}$, but both these conditions are clearly impossible. Therefore we can conclude that for all i and all j

$$(A\hat{y})_i = e_i \cdot A\hat{y} = \hat{x} \cdot A\hat{y} = c \qquad (\hat{x}^\mathsf{T} B)_j = \hat{x} \cdot Be_j = \hat{x} \cdot B\hat{y} = \tilde{c}$$

where $c, \tilde{c} \in \mathbb{R}$ are constants.

Remark. *We have just showed an equivalent statement to Lemma* 1.2 *in the lecture notes, but for the specific case of a Nash Equilibrium point contained in the interior of the state space. We can do better. If a Nash Equilibrium* $(\hat{x}, \hat{y})$ *is NOT in the interior of* $\Delta_A \times \Delta_B$ *then from the above proof we can conclude that*

$$\begin{cases} (A\hat{y})_i = c & \textit{whenever } \hat{x}_i \neq 0 \\ (\hat{x}^\mathsf{T} B)_j = \tilde{c} & \textit{whenever } \hat{y}_j \neq 0 \end{cases}$$

where $c, \tilde{c} \in \mathbb{R}$ *are constants. This gives us a complete reformulation of Lemma 1.2 for 2 player games.*

3) Consider the two player game encoded by the matrices

$$A = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix},$$

and let us choose to follow the second convention for 2 players games. We will firstly compute the Nash Equilibria of these matrices and then we will check if they are ESS. The Best Response maps are

$$\mathcal{BR}_A(y) = \arg\max_{x \in \Delta} x \cdot Ay = \arg\max_{x \in \Delta} 2 - (1 - y_1)x_1 = \begin{cases} \{e_2\} & \text{if } y \neq e_1 \\ \Delta & \text{if } y = e_1 \end{cases}$$

$$\mathcal{BR}_B(x) = \arg\max_{y \in \Delta} x \cdot By = \arg\max_{y \in \Delta} 2 - x_2 y_1 = \begin{cases} \{e_2\} & \text{if } x \neq e_1 \\ \Delta & \text{if } x = e_1 \end{cases}$$

from which we can read that (e_1, e_1) and (e_2, e_2) are Nash Equilibria: $e_1 \in \Delta = \mathcal{BR}_A(e_1)$ and $e_1 \in \Delta = \mathcal{BR}_B(e_1)$ and also $e_2 \in \mathcal{BR}_A(e_2)$ and $e_2 \in \mathcal{BR}_B(e_2)$. Note that (e_2, e_1) is an NE for this game as well. Since e_1 corresponds to strategy i, and e_2 to strategy ii, we can conclude that the strategies (i, i) and (ii, ii) are NE.

Next we want to see if such strategies are Evolutionary Stable. Recall the definition: $(\hat{x}, \hat{y})$ is an ESS if for all $\varepsilon > 0$ and all $(x, y) \in (\Delta_A \setminus \{\hat{x}\}) \times (\Delta_B \setminus \{\hat{y}\})$ then

$$x \cdot A(\varepsilon y + (1 - \varepsilon)\hat{y}) < \hat{x} \cdot A(\varepsilon y + (1 - \varepsilon)\hat{y})$$
$$(\varepsilon x + (1 - \varepsilon)\hat{x}) \cdot By < (\varepsilon x + (1 - \varepsilon)\hat{x}) \cdot B\hat{y}.$$

Firstly we will show that $(\hat{x}, \hat{y}) = (e_1, e_1)$ is NOT an ESS. Fix $\varepsilon > 0$, and take the point $\left(\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\right)$ then this choice yields

$$
\begin{aligned}
x \cdot A(\varepsilon y + (1-\varepsilon)\hat{y}) &= \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 - \frac{\varepsilon}{2} \\ \frac{\varepsilon}{2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 2 - \frac{\varepsilon}{2} \\ 2 \end{pmatrix} \\
&= 2 - \frac{\varepsilon}{4} > 2 - \frac{\varepsilon}{2} \\
&= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 2 - \frac{\varepsilon}{2} \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 - \frac{\varepsilon}{2} \\ \frac{\varepsilon}{2} \end{pmatrix} = \hat{x} \cdot A(\varepsilon y + (1-\varepsilon)\hat{y})
\end{aligned}
$$

which means that (e_1, e_1) is not an ESS.

On the other hand, the Nash Equilibrium $(\hat{x}, \hat{y}) = (e_2, e_2)$ is an ESS. Let $\varepsilon \in (0,1)$, and take any $(x, y) \in (\Delta_A \setminus \{e_2\}) \times (\Delta_B \setminus \{e_2\})$ then

$$
\begin{aligned}
x \cdot A(\varepsilon y + (1-\varepsilon)\hat{y}) &= \begin{pmatrix} x_1 \\ 1 - x_1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \varepsilon y_1 \\ 1 - \varepsilon y_1 \end{pmatrix} \\
&= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} 1 + \varepsilon y_1 \\ 2 \end{pmatrix} \\
&= x_1(1 + \varepsilon y_1) + 2(1 - x_1) \\
&< 2x_1 + 2(1 - x_1) = 2 \\
&= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 + \varepsilon y_1 \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \varepsilon y_1 \\ 1 - \varepsilon y_1 \end{pmatrix} = \hat{x} \cdot A(\varepsilon y + (1-\varepsilon)\hat{y})
\end{aligned}
$$

and

$$
\begin{aligned}
(\varepsilon x + (1-\varepsilon)\hat{x}) \cdot By &= \begin{pmatrix} \varepsilon x_1 \\ 1 - \varepsilon x_1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ 1 - y_1 \end{pmatrix} \\
&= 2\varepsilon x_1 + (2 - y_1)(1 - \varepsilon x_1) \\
&< 2\varepsilon x_1 + 2(1 - \varepsilon x_1) = 2 \\
&= \begin{pmatrix} \varepsilon x_1 \\ 1 - \varepsilon x_1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \hat{y} \cdot B(\varepsilon x + (1-\varepsilon)\hat{x})
\end{aligned}
$$

which confirms that (e_2, e_2), or (ii, ii) is an ESS.

2.2 Two players replicator dynamics

Exercise 2.2:
Consider the two players game given by the matrix

$$G = \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (-1,1) \end{pmatrix}$$

then, as we have done in the previous question, we can retrieve the two matrices defining this game

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad B = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Using Equations (17) from Section 2.2 of the notes we can write down the replicator equations of this game as

$$\begin{aligned} \dot{x_i} &= x_i((Ay)_i - x \cdot Ay) \\ &= x_i((-1)^{i+1} y_i - x_1 y_1 + x_2 y_2) \\ \dot{y_j} &= y_j((x^{\mathsf{T}} B)_j - x \cdot By) \\ &= y_j((-1)^j x_j + x_1 y_1 - x_2 y_2) \end{aligned}$$

where $x, y \in \Delta = \langle e_1, e_2 \rangle$, and $i, j \in \{1, 2\}$. Because of the specific shape of our phase space Δ we can rewrite these two equations using the fact that $x_2 = 1 - x_1$ and $y_2 = 1 - y_1$ for $x_1, y_1 \in [0,1]$

$$\begin{aligned} \dot{x_1} &= x_1(1 - x_1) \\ \dot{y_1} &= -y_1(1 - y_1). \end{aligned}$$

We can then proceed with the usual phase diagram analysis, as for any system of ODEs. The phase space of this system is the unit square $I^2 = [0,1] \times [0,1] \subset \mathbb{R}^2$. From now on we will drop the indexes in the equations. The derivative of x is zero along $\{x = 0\}$ and $\{x = 1\}$, whereas $\dot{y}$ is zero along $\{y = 0\}$ and $\{y = 1\}$. The four vertices of I^2 are singularities for the flow ($\dot{x} = \dot{y} = 0$). Along $\{0\} \times (0,1)$ and $\{1\} \times (0,1)$ we have $\dot{y} < 0$, and along $(0,1) \times \{0\}$ and $(0,1) \times \{1\}$, we have $\dot{x} > 0$. We can therefore conclude that $(0,1)$ is a sink (in black in the figure), $(0,0)$ and $(1,1)$ are saddles (in red in the figure), and $(1,0)$ is a sink (in white in the figure). The flow flows travels from close to $(0,1)$ towards $(1,0)$ without ever touching the boundary of the unit square. See Figure 13 for a sketch of the flow.

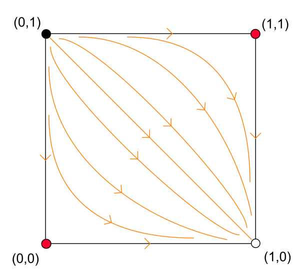


Figure 13: Flow for the game determined by G

2.3 Symmetric games

Exercise 2.3:
Consider a symmetric 2×2 game with payoff (square) matrices A and B, where $A = B^\mathsf{T}$. Suppose that $x(0) = y(0)$. We are going to show that if Player 1 plays strategy x against strategy y then that is equal to Player 2 playing x against y. We are going to work with the 2^{nd} convention for the replicator equation. Recall that if A, B are square matrices then $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$. Therefore

$$
\begin{aligned}
\dot{x_i} &= x_i((Ay)_i - x \cdot Ay) = x_i((B^\mathsf{T} y)_i - x \cdot B^\mathsf{T} y) \\
&= x_i((y^\mathsf{T} B)_i - x^\mathsf{T} B^\mathsf{T} y) = x_i((y^\mathsf{T} B)_i - y^\mathsf{T} Bx) \\
&= x_i((y^\mathsf{T} B)_i - y \cdot Bx) = \dot{y_i}.
\end{aligned}
$$

Under the assumption $x(0) = y(0)$ we can conclude that, by uniqueness of solutions of ODEs, that $x(t) = y(t)$ for all times. Recall that we have showed in part 3 of Exercise 1.3 that a solution always exist to every initial value problem (starting in Δ) and that such a solution exists for all times.

This question can be approached from a more geometric perspective. Consider the product space $\Delta \times \Delta$ and more specifically its diagonal

$$
D := \{(x, y) \in \Delta \times \Delta \mid x = y\}.
$$

We want to show that the space D is invariant under the action of $\dot{x} - \dot{y}$. Using Equations (18) from the lecture notes we can write

$$
\begin{aligned}
\dot{x} - \dot{y} &= x_i((Ay)_i - x \cdot Ay) - y_i((Ax)_i - x \cdot Ay) \\
&= (x_i - y_i)((A(x + y))_i - x \cdot Ay) - (x_i(Ax)_i - y_i(Ay)_i)
\end{aligned}
$$

which tells us that $\dot{x} - \dot{y}|_D = 0$. This means that the vector field $\dot{x} - \dot{y}$ is tangent to D at every point, hence there is no normal component pointing outwards from D. Therefore, D is invariant under $\dot{x} - \dot{y}$, and so if the flow $(x(t), y(t))$ starts on D, i.e. $x(0) = y(0)$, then $(x(t), y(t))$ is in D for all times t.

2.4 The 2×2 case

Exercise 2.4:
1) Consider the system of ODEs given by Equations (19) in Section 2.4 of the notes

$$
\begin{aligned}
\dot{x} &= x(1 - x)(\alpha_1 - y(\alpha_1 + \alpha_2)) \\
\dot{y} &= y(1 - y)(\beta_1 - x(\beta_1 + \beta_2)).
\end{aligned}
\tag{8}
$$

We want to understand all the different (non-degenerate) phase portraits that can arise from this system. Our state space is the usual unit square $I^2 = [0, 1] \times [0, 1]$ in $\mathbb{R}^2$, endowed with the subspace topology. Firstly, notice that along the boundary ∂I^2 at least one of the two derivatives is zero.

{x=0}	$\beta_1 > 0 \Rightarrow \dot{y} > 0$
{y=0}	$\alpha_1 > 0 \Rightarrow \dot{x} > 0$
{x=1}	$\beta_2 > 0 \Rightarrow \dot{y} < 0$
{y=1}	$\alpha_2 > 0 \Rightarrow \dot{x} < 0$

Table 3: Behaviour of Equation 8 along ∂I^2.

Table 3 summarises the behaviour of the flow restricted on the boundary. If we now look at the interior of the unit square we have two other lines along which one of the equations in 8 equals zero, namely $\{x = \frac{\beta_1}{\beta_1+\beta_2}\}$ along which $\dot{y} = 0$, and $\{y = \frac{\alpha_1}{\alpha_1+\alpha_2}\}$ along which $\dot{x} = 0$. Notice that the point $\theta = (\frac{\beta_1}{\beta_1+\beta_2}, \frac{\alpha_1}{\alpha_1+\alpha_2})$ is an equilibrium point for the system. For now we will assume that θ is in int I^2: this is equivalent to $\alpha_1\alpha_2 > 0$, and $\beta_1\beta_2 > 0$. In Figure 14 we have reported the direction of the flow when intersecting the nullclines $\{x = \frac{\beta_1}{\beta_1+\beta_2}\}$ (in red on the left) and $\{y = \frac{\alpha_1}{\alpha_1+\alpha_2}\}$ (in blue on the right).

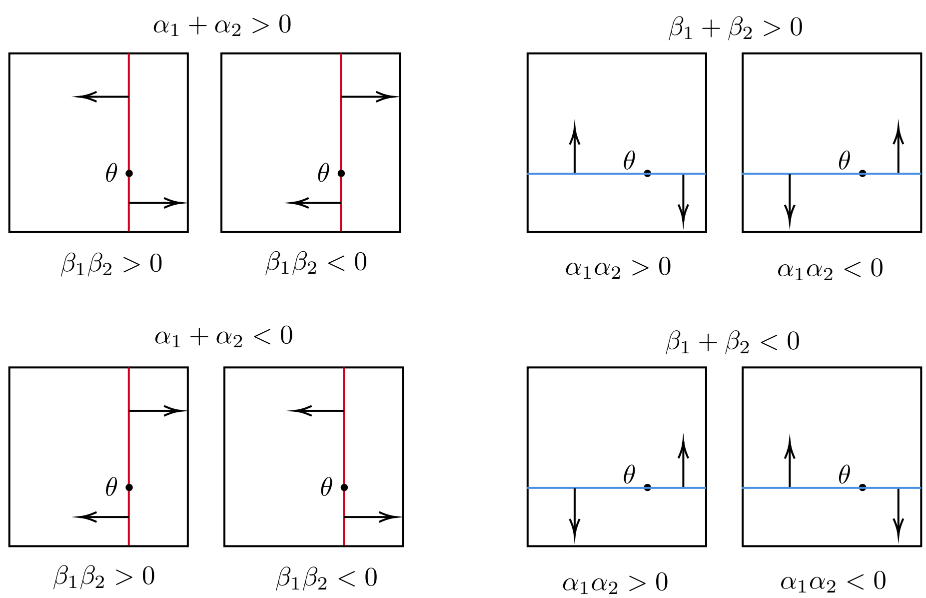


Figure 14: Possible flow directions along the nullclines in int I^2.

We can clearly see that the direction of the flow when crossing the nullclines only depends on $\alpha_1\alpha_2$ and $\beta_1\beta_2$. All the possible phase portraits that Equation 8 are described in Proposition 2.1 in the lecture notes. Given the assumption $\theta \in$ int I^2 we have made before, we are interested for now in understanding the portraits associated with case (i) and (iii). Set $\beta_1 > 0$ therefore fixing the direction of our flow (one gets the same portraits but with the flow direction reversed for $\beta_1 < 0$, as suggested in Figure 14). From Table 3 we know that if $\alpha_1\beta_1 > 0$ then the points $(0,0)$, and $(1,1)$ are sources, whilst $(0,1)$, and $(1,0)$ are sinks. On the other hand, $\alpha_1\beta_1 < 0$, translates to the flow travelling clockwise around ∂I^2. We can conclude that the left phase portrait in Figure 13 in the lecture notes corresponds to case (i), whereas the right one corresponds to (iii).

In Figure 16 you can see the direction of the derivatives in the various quadrants for the cases we have just discussed (under the underlying assumption $\beta_1 > 0$.

What happens if $\theta \notin \text{int } I^2$? If either $\alpha_1\alpha_2 < 0$ or $\beta_1\beta_2 < 0$ we obtain a *dominated strategy* type of system, case (ii) in Proposition 2.1. Assume, for the sake of discussion, that $\alpha_1 < 0$, and $\alpha_2 > -\alpha_2, \beta_1 > 0$, and $\beta_2 > 0$ (every other case is either similar or simpler). Then we have $\alpha_1\alpha_2 < 0$, $\beta_1\beta_2 > 0$, the nullcline $\{x = \frac{\beta_1}{\beta_1+\beta_2}\}$ is still in int I^2, so we obtain a phase diagram as in Figure 15. The yellow dot in the top right corner represents the dominating strategy, whereas the red dotted line is $\{x = \frac{\beta_1}{\beta_1+\beta_2}\}$. Picking different values for the α's and β's will surely change the dominating strategy, the direction of the flow, and the presence of nullclines, but the overall shape of the phase diagram will always be the same.

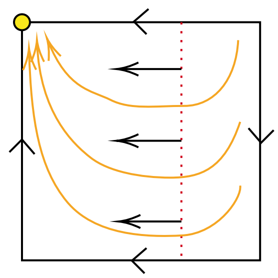
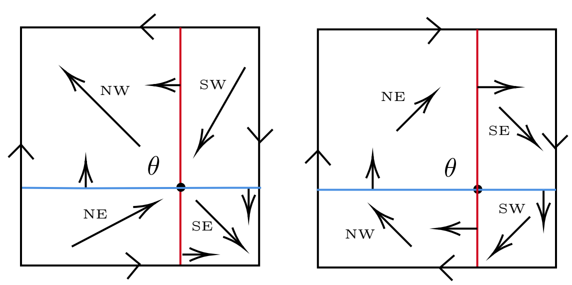


Figure 15: Dominated Strategy

A full explanation of the terminology can be found in *Hofbauer, Sigmund – Evolutionary games and population dynamics*. The term *dominated strategy* is illustrated in Section 8.3, whereas *battle of the sexes* or *coordination game* is explained in Section 10.2. The term *zero-sum case* includes not only zero-sum games, but all games in which the total payoff between players is zero, i.e. the net change of global wealth is zero.



Cooperation Game Zero-Sum Case

Figure 16: Direction of the flow for case (i) and (iii) assuming $\beta_1 > 0$.

We will now analyse some specific examples of games for every type of phase portrait described in Proposition 2.1. Please note we will adopt the second convention from now, and we will denote by $\Delta_A \times \Delta_B$ the total phase space (I^2 is a reparameterisation of such space). Let

$$A = \begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}$$

be the matrices describing the *stag hunt* game. Then

$$\alpha_1 = -1 \quad \alpha_2 = -1, \quad \beta_1 = -1, \quad \beta_2 = -1$$

hence $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, $\alpha_1\beta_1 > 0$: this is a coordination game. We have two pure dominating strategies, and a Mixed Nash Equilibrium in the interior of I^2. The Mixed Nash Equilibrium in I^2 is given by $\theta = \left(\frac{\beta_1}{\beta_1+\beta_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}\right)$ which corresponds to $\left(\left(\frac{\beta_1}{\beta_1+\beta_2}, \frac{\beta_2}{\beta_1+\beta_2}\right), \left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)\right) = \left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}\right)\right)$. The other Pure Nash Equilibria can be computed by looking at the best responses

$$\begin{aligned}
\mathcal{BR}_A(e_1) &= \arg\max_{x\in\Delta} x \cdot A\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) = \arg\max_{x\in\Delta} 4x_1 + 3x_2 = \{e_1\} \\
\mathcal{BR}_A(e_2) &= \arg\max_{x\in\Delta} x \cdot A\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right) = \arg\max_{x\in\Delta} x_1 + 2x_2 = \{e_2\} \\
\mathcal{BR}_B(e_1) &= \arg\max_{y\in\Delta} \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) \cdot By = \arg\max_{y\in\Delta} 4y_1 + 3y_2 = \{e_1\} \\
\mathcal{BR}_B(e_2) &= \arg\max_{y\in\Delta} \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right) \cdot By = \arg\max_{y\in\Delta} y_1 + 2y_2 = \{e_2\}
\end{aligned}$$

so (e_1, e_1) and (e_2, e_2) are the equilibria we were looking for, which corresponds to the strategies (C, C), and (D, D).

Another coordination game example is given by the *battle of sexes* game described by

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

The coefficients for this game are given by

$$\alpha_1 = -2, \qquad \alpha_2 = -3, \qquad \beta_1 = -3, \qquad \beta_2 = -2$$

which confirms that this is a coordination game since $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, and $\alpha_1\beta_1 > 0$. As before we have a Mixed Nash Strategy given by $\left(\left(\frac{3}{5}, \frac{2}{5}\right), \left(\frac{2}{5}, \frac{3}{5}\right)\right)$. The two Pure Nash Strategies are given again by (e_1, e_1) and (e_2, e_2).

Consider the classic *Prisoner's Dilemma* game described by the matrices

$$A = \begin{pmatrix} 2 & 0 \\ 3 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}$$

given in Example 0.3 in the lecture notes. For this game we have that

$$\alpha_1 = -1, \quad \alpha_2 = 1, \quad \beta_1 = -1, \quad \beta_2 = 1$$

which implies $\alpha_1\alpha_2 = -1 < 0$ and $\beta_1\beta_2 = -1 < 0$. The Prisoner's Dilemma falls under the dominated strategy category, therefore there is one dominating pure strategy given by the Nash Equilibrium (e_2, e_2). As a sanity check

$$\begin{aligned}
\mathcal{BR}_A(e_2) &= \arg\max_{y\in\Delta} y \cdot Ae_2 = \arg\max_{y\in\Delta} y_2 = \{e_2\} \\
\mathcal{BR}_B(e_2) &= \arg\max_{x\in\Delta} e_2 \cdot Bx = \arg\max_{x\in\Delta} x_2 = \{e_2\}
\end{aligned}$$

as we claimed.

Finally, let us look at zero sum type of games. The first we want to look at is described by

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad B = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

which give us coefficients

$$\alpha_1 = 1, \qquad \alpha_2 = -1, \qquad \beta_1 = -1, \qquad \beta_2 = 1.$$

Since $\alpha_1\alpha_2 < 0$, and $\beta_1\beta_2 < 0$ then we have a dominated strategy type of game. As before, we only have one Pure Nash Equilibrium

$$\mathcal{BR}_A(e_2) = \arg\max_{y\in\Delta} y \cdot Ae_2 = \arg\max_{y\in\Delta} -y_2 = \{e_1\}$$
$$\mathcal{BR}_B(e_1) = \arg\max_{x\in\Delta} e_1 \cdot Bx = \arg\max_{x\in\Delta} -x_1 = \{e_2\}$$

given by (e_1, e_2).

The last game we want to analyse is described by

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

which give us coefficients

$$\alpha_1 = -1, \qquad \alpha_2 = -1, \qquad \beta_1 = 1, \qquad \beta_2 = 1.$$

Since $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, and $\alpha_1\beta_1 < 0$ then we have a zero sum game with interior Nash Equilibrium, given once again by $\left(\left(\frac{\beta_1}{\beta_1+\beta_2}, \frac{\beta_2}{\beta_1+\beta_2}\right), \left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)\right) = \left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}\right)\right)$. We will see in the next section that for this game the solution to this specific type of game is described by simple close periodic orbits (topological circles). This means that θ is not asymptotically stable (only Lyapunov stable).

2) Recall that we proved in part 2 of Exercise 1.3 that adding constant column vectors to a matrix does not change its replicator dynamics. Therefore, it is possible that two different matrices induce a phase portrait belonging to the category of zero sum games.

3) As hinted in the question, we will firstly linearise our system around the equilibrium $\theta = \left(\frac{\beta_1}{\beta_1+\beta_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}\right)$. Consider the perturbed point $\tilde{\theta} = \left(\frac{\beta_1}{\beta_1+\beta_2} + \varepsilon_1, \frac{\alpha_1}{\alpha_1+\alpha_2} + \varepsilon_2\right)$ for $\varepsilon_1, \varepsilon_2 > 0$ small, then the equations for the replicator dynamics will give us

$$\dot{\varepsilon}_1 = -\frac{\beta_1\beta_2(\alpha_1+\alpha_2)}{(\beta_1+\beta_2)^2}\varepsilon_2 + \mathcal{O}(\varepsilon^2)$$
$$\dot{\varepsilon}_2 = -\frac{\alpha_1\alpha_2(\beta_1+\beta_2)}{(\alpha_1+\alpha_2)^2}\varepsilon_1 + \mathcal{O}(\varepsilon^2)$$

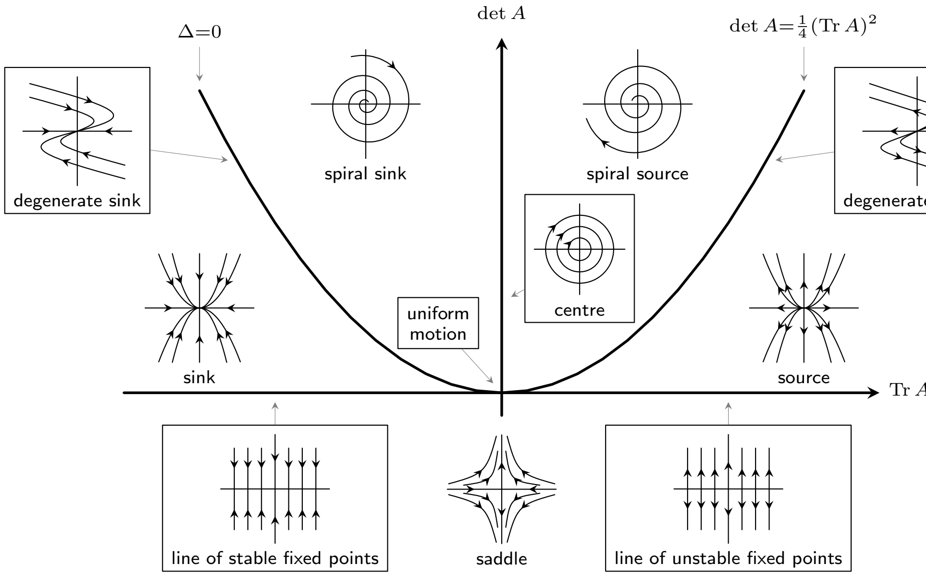


Figure 17: Classification of Phase Portraits in the (det A, Tr A)-plane [1]

which leads us to the linearisation matrix at θ

$$L = \begin{pmatrix} 0 & -\frac{\beta_1\beta_2(\alpha_1+\alpha_2)}{(\beta_1+\beta_2)^2} \\ -\frac{\alpha_1\alpha_2(\beta_1+\beta_2)}{(\alpha_1+\alpha_2)^2} & 0 \end{pmatrix}.$$

This matrix has a very special shape since its trace, $\mathrm{Tr}\,L$, is always zero, which severely restricts the possible flow behaviour near θ.

By looking at the Poincaré Diagram in Figure 17 we see that there are only two possibilities for the flow close to θ (excluding the degenerate case): either $\det L$ is positive and we have that the flow generates concentric ellipses (purely imaginary eigenvalues), or $\det L$ is negative and θ is a saddle point (the real parts of the two eigenvalues have opposite sign).

To start with, the determinant of L is given by

$$\det L = -\frac{\alpha_1\alpha_2\beta_1\beta_2}{(\alpha_1+\alpha_2)(\beta_1+\beta_2)}$$

therefore the conditions of Proposition 2.1 will uniquely determine the positivity of it.

For zero sum case with interior Nash Equilibrium $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, and $\alpha_1\beta_1 > 0$, which means that either α_1, α_2 are both positive, and β_1, β_2 are negative, or vice-versa. In both cases $\det L$ is positive (beware of the minus sign

[1] Adapted from: https://tex.stackexchange.com/questions/347201/

in front of it), and so θ is the centre of concentric ellipses (the ratio between the sizes of minor and major axes of the ellipses depends on the ration of the modulus of the imaginary eigenvalues).

For coordination games we know that $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, and $\alpha_1\beta_1 > 0$, which means that $\alpha_1, \alpha_2, \beta_1, \beta_2$ have all the same sign. This is sufficient to show that $\det L$ is negative and hence θ is a saddle point.

We will now directly show that the orbits of the replicator dynamics circle around θ. Consider the Lyapunov function

$$P(x,y) = x^{-\beta_1}(1-x)^{-\beta_2}y^{\alpha_1}(1-y)^{\alpha_2},$$

and we claim that this function is constant along the solutions of the replicator dynamics whenever $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, and $\alpha_1\beta_1 > 0$. Notice that P is always positive in I^2, and vanishes on its boundary. If we compute the time logarithmic derivative of P

$$\begin{aligned}\frac{\dot{P}}{P}(x,y) = \dot{\log P} = {} & \frac{d}{dt}(-\beta_1 \log x - \beta_2 \log(1-x) + \alpha_1 \log y + \alpha_2 \log(1-y)) \\ = {} & -\beta_1\frac{\dot{x}}{x} + \beta_2\frac{\dot{x}}{1-x} + \alpha_1\frac{\dot{y}}{y} - \alpha_2\frac{\dot{y}}{1-y} \\ = {} & -(\alpha_1 - y(\alpha_1+\alpha_2))(\beta_1 - \beta_1 x - \beta_2 x) + (\beta_1 - x(\beta_1+\beta_2))(\alpha_1 - \alpha_1 y - \alpha_2 y) \\ = {} & -(\alpha_1 - y(\alpha_1+\alpha_2))(\beta_1 - x(\beta_1+\beta_2)) + (\alpha_1 - y(\alpha_1+\alpha_2))(\beta_1 - y(\beta_1+\beta_2)) \\ = {} & 0\end{aligned}$$

we see that $\dot{P} = 0$ along the orbits of the solution to Equation 8.

This means that P is constant along orbits, therefore the orbits of Equation 8 are level sets of P.

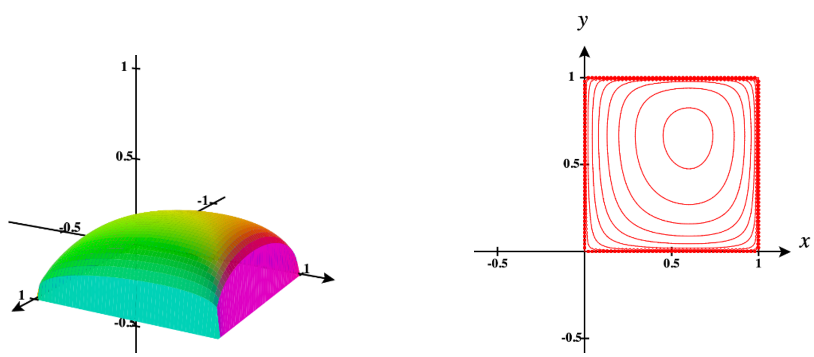


Figure 18: 3D and contour plot of the function P for $\alpha_1 = \frac{1}{3}, \alpha_2 = \frac{1}{6}, \beta_1 = -\frac{1}{2}, \beta_2 = -\frac{1}{3}$.

In Figure 18 we can see a contour plot for the function P for some arbitrarily chosen values of $\alpha_1, \alpha_2, \beta_1$, and β_2. We can see that the level sets of P are closed simple lines, or topological circles. Those represent the shape of the orbits of the flow induced by Equation 8 in I^2.

4) All the hard work is now done. We have showed, by computing the linearization matrix at θ that the equilibrium point can only either be a saddle (unstable) or orbits cycle around it as in Figure 18 (Lyapunov stable). As we proved in 1.1 in the lecture notes, Evolutionary Stable Strategy are asymptotically stable equilibria for the replicator dynamics, but since θ in our case is either unstable or Lyapunov stable (this is weaker than asymptotically stable), then we can conclude that these games admit no ESS in the interior of I^2.

2.5 A 3×3 replicator dynamics systems with chaos

Exercise 2.5:
1) Fix $\varepsilon \in (0,1)$ (in order to simplify the calculations), and the two matrices

$$A = \begin{pmatrix} \varepsilon & -1 & 1 \\ 1 & \varepsilon & -1 \\ -1 & 1 & \varepsilon \end{pmatrix} \qquad B = \begin{pmatrix} -\varepsilon & -1 & 1 \\ 1 & -\varepsilon & -1 \\ -1 & 1 & -\varepsilon \end{pmatrix}$$

which describe a 3×3 game with two players. We wish to compute the Nash Equilibria of such a game. In order to maintain consistency with the lecture notes we will adopt the first convention. The state space will be denoted by $\Delta_A \times \Delta_B$. As usual, we will firstly show that we have only one Nash Equilibrium in the interior of $\Delta_A \times \Delta_B$, and then we will move to the boundary. Please note that we will freely use the letters i, j, k to denote indices, these have to be understood as all different elements of $\mathbb{Z}/3\mathbb{Z}$.

Let us consider the indifference lines in Δ_A

$$\begin{aligned} Z^A_{1,2} &= \{(Ay)_1 = (Ay)_2\} = \{(3-\varepsilon)y_1 + (3+\varepsilon)y_2 = 2\} \\ Z^A_{2,3} &= \{(Ay)_2 = (Ay)_3\} = \{(3-\varepsilon)y_2 + (3+\varepsilon)y_3 = 2\} \\ Z^A_{1,3} &= \{(Ay)_1 = (Ay)_3\} = \{(3+\varepsilon)y_1 + (3-\varepsilon)y_3 = 2\} \end{aligned}$$

and in the simplex Δ_B

$$\begin{aligned} Z^B_{1,2} &= \{(Bx)_1 = (Bx)_2\} = \{(3+\varepsilon)x_1 + (3-\varepsilon)x_2 = 2\} \\ Z^B_{2,3} &= \{(Bx)_2 = (Bx)_3\} = \{(3+\varepsilon)x_2 + (3-\varepsilon)x_3 = 2\} \\ Z^B_{1,3} &= \{(Ax)_1 = (Ax)_3\} = \{(3-\varepsilon)x_1 + (3+\varepsilon)x_3 = 2\}. \end{aligned}$$

In Figure 19 we reported all the indifference lines we have just computed, together with the best response for every convex region. To estimate the best response in each of these region it is enough to compute the best response at each corner which accounts to

$$\mathcal{BR}_A(e_i) = \{e_{i+1}\}$$

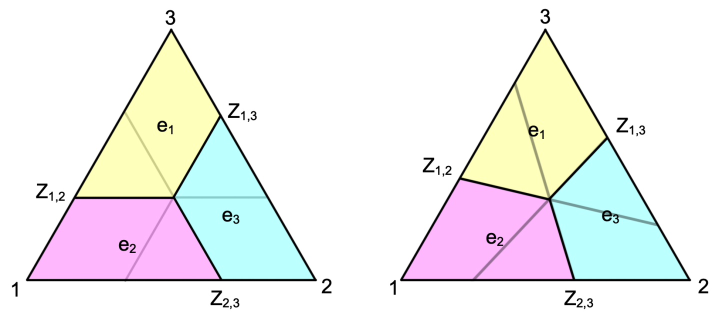


Figure 19: Indifference lines and best response in Δ_A (left), and in Δ_B (right).

and

$$\mathcal{BR}_B(e_i) = \{e_{i+1}\}.$$

Along every indifference line $Z^A_{i,j}$ there is a portion along which $(Ay)_k \leq (Ay)_i = (Ay)_j$ (denoted by a black segment), and a portion along which the opposite inequality holds (denoted by a light grey segment). The same notation has been adopted in Δ_B. Henceforth, when we will say "indifference line" will refer to the black segment of that indifferent line, i.e. we abuse notation and redefine $Z^A_{i,j} \coloneqq Z^A_{i,j} \cap \{(Ay)_k \leq (Ay)_i = (Ay)_j\}$, and $Z^B_{i,j} \coloneqq Z^B_{i,j} \cap \{(Bx)_k \leq (Bx)_i = (Ay)_j\}$. As you will see in Chapter 4, the indifference lines are discontinuity lines for the best response map, hence why we only worried about computing the best response at each corner.

The indifference lines, in both simplices, meet at the point $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. This means that the point $((\frac{1}{3}, \frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}))$ is an interior Nash Equilibrium for this system. This clearly is the only internal Nash Equilibrium, since the indifference lines do not intersect again.

The last thing we are left with is to check for Nash Equilibria along the boundary of our space. These points can only appear as intersections of indifference lines and a side. Notice that $Z^A_{i,j}$ intersects the side $\langle e_{i-1}, e_{j-1}\rangle$, and the best response at the intersection point is given by $\mathcal{BR}_A(Z^A_{i,j} \cap \langle e_{i-1}, e_{j-1}\rangle) = \langle e_i, e_j\rangle$. We find an identical picture in Δ_B, meaning that $\mathcal{BR}_B(Z^B_{i,j} \cap \langle e_{i-1}, e_{j-1}\rangle) = \langle e_i, e_j\rangle$. Therefore

$$\begin{aligned}
\mathcal{BR}_A(Z^A_{i,j} \cap \langle e_{i-1}, e_{j-1}\rangle) &= \langle e_i, e_j\rangle \\
\mathcal{BR}_B(Z^B_{i+1,j+1} \cap \langle e_i, e_j\rangle) &= \langle e_{i+1}, e_{j+1}\rangle
\end{aligned}$$

but since the intersection between indifference lines and sides does not happen at the corners of the simplices (remember $\varepsilon \in (0, 1)$), we can conclude we have no Nash Equilibria on the boundary of the state space. The only Nash Equilibrium for this game is given by

$$\left(\begin{pmatrix}\frac{1}{3}\\ \frac{1}{3}\\ \frac{1}{3}\end{pmatrix}, \begin{pmatrix}\frac{1}{3}\\ \frac{1}{3}\\ \frac{1}{3}\end{pmatrix}\right).$$

2) We now want to understand the flow described by the replicator dynamics induced by the matrices A and B. More specifically, we want to study the direction of the flow along the edges connecting the vertices of the space $\Delta_A \times \Delta_B$, and show that it corresponds to the one represented in Figure 14 in the Lecture Notes. Firstly, recall that we are considering the system of differential equations

$$\begin{aligned}\dot{x}_i &= x_i((Ay)_i - x \cdot Ay)\\ \dot{y}_j &= y_j((Bx)_j - y \cdot Bx)\end{aligned}$$

and in order to understand the direction of the flow we want to look at the following ratios

$$\begin{aligned}\left(\frac{x_i}{x_j}\right)' &= \frac{\dot{x}_i x_j - x_i \dot{x}_j}{x_j^2} = \frac{x_i}{x_j}\left[(Ay)_i - (Ay)_j\right]\\ \left(\frac{y_i}{y_j}\right)' &= \frac{\dot{y}_i y_j - y_i \dot{y}_j}{y_j^2} = \frac{y_i}{y_j}\left[(Bx)_i - (Bx)_j\right]\end{aligned}$$

where x will always denote an element from Δ_A, and y an element from Δ_B, and $i \neq j$. We will now proceed to calculate a few of these ratios. Recall that we have that R is associated to e_1, P to e_2, and S to e_3. Let us say that we want to understand the direction of the flow between the points $(P, P) = (e_2, e_2)$ and $(P, S) = (e_2, e_3)$, then this means that we are interested in the sign of $(y_2/y_3)'$ along the constraint $x = e_2$ (notice sign $(y_2/y_3)' = -\text{sign } (y_3/y_2)'$). Hence for $y \in \langle e_2, e_3\rangle \setminus \{e_2, e_3\}$

$$\left(\frac{y_2}{y_3}\right)'\Bigg|_{x=e_2} = \frac{y_2}{y_3}((Be_2)_2 - (Be_2)_3) = -(1+\varepsilon)\frac{y_2}{y_3} < 0$$

which translates to the flow moving from (e_2, e_2) towards (e_2, e_3) in $\Delta_A \times \Delta_B$. This means that the flow goes from (P, P) to (P, S).

Similarly we can show that the flow goes from (R, P) to (P, P). In order to see this let us compute for $x \in \langle e_1, e_2\rangle \setminus \{e_1, e_2\}$, and $y = e_2$

$$\left(\frac{x_1}{x_2}\right)'\Bigg|_{y=e_2} = \frac{x_1}{x_2}\left[(Ae_2)_1 - (Ae_2)_2\right] = -(1+\varepsilon)\frac{x_1}{x_2} < 0$$

which confirms that the flow goes from (R, P) to (P, P). Similar calculations give us the direction of the flow along all the edges of the graph in Figure 14.

3) The space $\Delta_A \times \Delta_B$ is 6 dimensional, and it can be reduced to 4 dimensions using the definition of simplex, for example by discarding the 3$^{\text{rd}}$ component of

the vectors $x \in \Delta_A$, and $y \in \Delta_B$ since $x_3 = 1 - x_1 - x_2$, and $y_3 = 1 - y_1 - y_2$. Even if we reduce $\Delta_A \times \Delta_B$ to a 4 dimensional object, in order to represent it on a sheet of paper we need to project it into $\mathbb{R}^2$ or $\mathbb{R}^3$. One of the ways to represent $\Delta_A \times \Delta_B$ in $\mathbb{R}^2$ can be seen in Figure 14 in the Lecture Notes. In order to replicate such a graph we need to find the right projection matrix. Recall that Rock corresponds to e_1, Paper to e_2, and Scissors to e_3. Henceforth, we will work under the identification $\left(\left(\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right), \left(\begin{smallmatrix} y_1 \\ y_2 \\ y_3 \end{smallmatrix}\right)\right) = (x_1, x_2, x_3, y_1, y_2, y_3)^\mathsf{T}$.

The matrix

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

induces a linear map from $\mathbb{R}^6$ to $\mathbb{R}^2$. The projection induced by X is too restrictive since we loose too much information: consider $(R, P) = (1, 0, 0, 0, 1, 0)$ and $(R, S) = (1, 0, 0, 0, 0, 1)$, then

$$X(R, P) = X(R, S) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

More specifically, $X(R, \cdot) = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, $X(P, \cdot) = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$, and $X(S, \cdot) = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, independently from the last three entries of the vectors, i.e. independently from the component coming from Δ_B.

In order to obtain a more useful and meaningful projection of $\Delta_A \times \Delta_B$ we aim at a matrix similar to X, but under which the nine points $(R, R), (R, P), (R, S), (P, R), (P, P), (P, S), (S, R), (S, P), (S, S)$ have all distinct images. For example, consider

$$X = \begin{pmatrix} 3.65 & -1.35 & 1.35 & 5.35 & 1.35 & 1.45 \\ 0.40 & 0.40 & 4.60 & 1.90 & -0.40 & 4.40 \end{pmatrix}$$

then this give us exactly what we want

$$\begin{array}{lll} X(R, R) = (9.00, 2.30) & X(R, P) = (5.00, 0.00) & X(R, S) = (5.10, 4.80) \\ X(P, R) = (4.00, 2.30) & X(P, P) = (0.00, 0.00) & X(P, S) = (0.10, 4.80) \\ X(S, R) = (6.70, 6.50) & X(S, P) = (2.70, 4.20) & X(S, S) = (2.80, 9.00). \end{array}$$

Notice this matrix was computed to generate the diagram in Figure 14 in the Lecture notes.

4) See the appendix of the lecture notes.

3 Iterated Prisoner Dilemma (IRP) and the Role of Reciprocity

3.1 Repeated games with unknown time length

Exercise 3.1:
Recall that the bimatrix for the Prisoner's Dilemma is given by

$$\begin{pmatrix} (-1,-1) & (-3,0) \\ (0,-3) & (-2,-2) \end{pmatrix}$$

where $(-2,-2)$ is the payoff if both prisoners defect. As we have seen in the notes the Nash Equilibrium of this game, given by both parties always defecting, is rather sub-optimal since it leads to a fairly poor payoff. We will now investigate how playing different strategies (namely how cooperating) can usually increase the total payoff if a game is played for a long time (as $t \to \infty$).

Now suppose that Player 1 defects with probability $p \in [0,1]$ and Player 2 with probability $q \in [0,1]$. All the computations will be carried out for Player 1, but given the symmetry of the game the whole discussion immediately extends to Player 2. The payoff at round n is given by

$$\begin{aligned} A_n =&(-2)pq + (-1)(1-p)(1-q) + (-3)(1-p)q+ \\ &+ (0)p(1-q) = p - 2q - 1. \end{aligned}$$

Therefore the total payoff, as explained in the notes, is given by

$$A(\omega) = \sum_{i=1}^{\infty} A_i \omega^{i-1} = \frac{p-2q-1}{1-\omega}$$

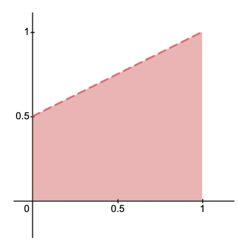


Figure 20: Probability square for (p,q)

where $\omega \in (0,1)$ is the chance of play the following turn. The expected payoff is therefore given by

$$\mathbb{E}(\text{Payoff}_A) = \frac{A(\omega)}{\frac{1}{1-\omega}} = p - 2q - 1.$$

For $p = 1$, and $q = 1$ we have that $\mathbb{E}(\text{Payoff}_A) = -2$, which precisely corresponds to the expected payoff if both Players always defect. In Figure 20 we have the probability unit square $I^2 = [0,1] \times [0,1]$ with p along the x–axis, and q along the y–axis. The red shaded region represents all the tuples (p,q) which give an expected payoff to Player 1 greater than -2. The shaded region accounts for $3/4$ of the total area, meaning that if they choose to sometime cooperate ($p < 1$) then it is likely that they will get a higher payoff (in the long run) that playing always defect.

3.2 The three strategies AllC, AllD, TFT

Exercise 3.2:
Consider the standard donation game. In Section 3.1 we have analysed the various payoffs corresponding to the strategies always defect (AllD), always cooperate (allC), tit for tat (TFT). Let us consider a new strategy TFTT: a player defects only when the other player defects twice. We want to establish a payoff matrix as Matrix (21) in Section 3.2. As before, $\omega \in (0,1)$ represents the probability of playing a new round. We will assume that for TFT, and TFTT strategy the player will start by cooperating.

If Player 1 plays $TFTT$ then whenever Player 2 plays AllC, TFT, or TFTT Player 1's payoff is given by $\frac{b-c}{1-\omega}$, since both players are constantly cooperating. The interesting case is whenever Player 2 plays AllD. In this case we have that Player 1's payoff is given by

$$A_1 = -c, \quad A_2 = -c \quad A_n = 0, \quad \text{for } n \geq 3$$

so that $A(\omega) = -c(1+\omega)$.

Symmetrically if Player 1 plays AllD against TFTT, then we have that their payoff is given by

$$A_1 = b, \quad A_2 = b, \quad A_n = 0 \quad \text{for } n \geq 3$$

so that $A(\omega) = b(1+\omega)$.

The payoff matrix is given by

$$\frac{1}{1-\omega}\begin{pmatrix} b-c & -c & b-c & b-c \\ b & 0 & b(1-\omega) & b\frac{1+\omega}{1-\omega} \\ b-c & -c(1-\omega) & b-c & b-c \\ b-c & -c\frac{1+\omega}{1-\omega} & b-c & b-c \end{pmatrix}$$

where the fourth row represents Player 1 playing TFTT, and the fourth column represents Player 2 playing TFTT.

3.3 The replicator dynamics associated to a repeated game with the AllC, AllD, TFT strategies

Exercise 3.3:
We want to calculate the Evolutionary Stable Strategies and Nash Equilibria of the matrix

$$A = \begin{pmatrix} -c & -c & b\omega - c \\ 0 & 0 & 0 \\ -c & -c(1-\omega) & b\omega - c \end{pmatrix}.$$

We will assume that $\omega \in (0,1)$. Notice that the matrix we just wrote has the same Evolutionary Stable Strategies and Nash Equilibria as Matrix (21) in Section 3.1.

There are two cases we have to consider. Firstly assume $b\omega < c$. When calculating the best response to a strategy $x \in \Delta$ we have

$$\mathcal{BR}(x) = \arg\max_{y\in\Delta} y \cdot Ax = \arg\max_{y\in\Delta}(b\omega x_3 - c)y_1 + (b\omega x_3 - c + \omega c x_2)y_3.$$

Thanks to the assumptions $b\omega < c$ and $\omega < 1$ it follows that

- $b\omega x_3 - c < cx_3 - c = -c(1 - x_3) \leq 0$;
- $b\omega x_3 - c + \omega c x_2 < cx_3 - c + cx_2 = -c(1 - x_1 - x_2) \leq 0$

which means that both the coefficients we are trying to minimise are negative, therefore

$$\mathcal{BR}(x) = \{e_2\} \qquad \text{for all } x \in \Delta.$$

The only Nash Equilibrium in this case is e_2, but it actually is a strict Nash Equilibrium

$$x \cdot Ae_2 = -cx_1 + (\omega c - c)x_3 < -cx_1 \leq 0 = e_2 \cdot Ae_2$$

where $x \in \Delta \setminus \{e_2\}$. Therefore, if $b\omega < c$ and $\omega < 1$ then e_2 is a strict Nash Equilibrium, therefore an ESS and Nash Equilibrium.

Now we will assume $b\omega > c$. As we have seen before we have for $x \in \Delta$

$$Ax = \begin{pmatrix} b\omega x_3 - c \\ 0 \\ b\omega x_3 - c + c\omega x_2 \end{pmatrix} = \begin{pmatrix} (Ax)_1 \\ 0 \\ (Ax)_1 + c\omega x_2 \end{pmatrix}.$$

We can compute the indifference lines

$$\begin{aligned} Z_{1,2} &= \{x_3 = \frac{c}{b\omega}\} \\ Z_{2,3} &= \{cx_2 + bx_3 = \frac{c}{\omega}\} \\ Z_{1,3} &= \{x_2 = 0\} = \langle e_1, e_3 \rangle, \end{aligned}$$

which intersect at the point $\tilde{q} = \left(\frac{b\omega - c}{b\omega}, 0, \frac{c}{b\omega}\right)$, which is automatically a Nash Equilibrium. Clearly, there are no NE in the interior of Δ.

Let us compute the best response at the various corners of the simplex. We obtain

$$\begin{aligned} \mathcal{BR}(e_1) &= \arg\max_{y\in\Delta} y \cdot Ae_1 \\ &= \arg\max_{y\in\Delta} -c(y_1 + y_3) = \{e_2\} \\ \mathcal{BR}(e_2) &= \arg\max_{y\in\Delta} y \cdot Ae_2 \\ &= \arg\max_{y\in\Delta} -cy_1 - c(1-\omega)y_3 = \{e_2\} \qquad \text{since } \omega < 1 \\ \mathcal{BR}(e_3) &= \arg\max_{y\in\Delta} y \cdot Ae_3 \\ &= \arg\max_{y\in\Delta}(b\omega - c)(y_1 + y_3) = \langle e_1, e_3 \rangle \quad \text{since } b\omega - c > 0 \end{aligned}$$

which tell us that e_2, e_3 are Nash Equilibria.

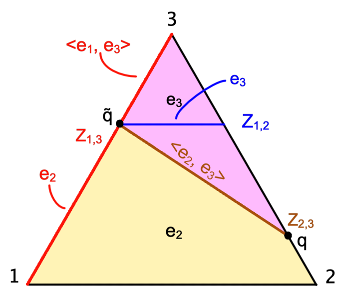


Figure 21: Indifference lines and $\mathcal{BR}$ when $b\omega > c > 0$.

Similar calculations tell us the best response along the indifference lines, which we have reported in Figure 21. From it we can read off the remaining Nash Equilibria. If $x \in Z_{1,3} = \langle e_1, e_3 \rangle$, and $x_3 > \frac{c}{b\omega}$, then the best response is given by $\langle e_1, e_3 \rangle$, which implies we have a line of NE, between $\tilde{q}$, and e_3 (included). The last candidate as a Nash Equilibrium is given by $q = \left(0, \frac{b\omega - c}{\omega(b-c)}, \frac{c(1-\omega)}{\omega(b-c)}\right)$, which corresponds to the intersection between $Z_{2,3}$ and the side $\langle e_2, e_3 \rangle$. Along the indifference line $Z_{2,3}$ the Best Response is $\langle e_2, e_3 \rangle$ (except at $\tilde{q}$, hence q is a NE.

The Nash Equilibria of this system, denoted NE, are

$$\text{NE} = \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{b\omega - c}{(b-c)\omega} \\ \frac{c(1-\omega)}{(b-c)\omega} \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 1 - p_3 \\ 0 \\ p_3 \end{pmatrix} \right\}_{p_3 \in [\frac{c}{b\omega}, 1]},$$

and in order to simplify notation we will refer to the last family of vectors as Γ, so that $\text{NE} = \{e_1, q\} \cup \Gamma$.

We are left with showing which of these points are ESS. As before, e_2 is an ESS since by Lemma 1.3 in the notes, we need to show that for y close to e_2 we have that $y \cdot Ay < e_2 \cdot Ay = 0$. So if we let $0 \leq \delta, \tau$ and $0 < \varepsilon$, where $\delta + \tau = \varepsilon$, and $y = \left(\begin{smallmatrix} \delta \\ 1-\varepsilon \\ \tau \end{smallmatrix} \right)$ then

$$\begin{aligned} y \cdot Ay &= \begin{pmatrix} \delta \\ 1 - \varepsilon \\ \tau \end{pmatrix} \cdot \begin{pmatrix} b\omega\tau - c \\ 0 \\ b\omega\tau - c + c\omega - c\omega\varepsilon \end{pmatrix} = \\ &= -c\delta - c\tau + c\omega\tau + \mathcal{O}(\varepsilon^2) \\ &\leq -c\varepsilon(1-\omega) + \mathcal{O}(\varepsilon^2) < 0 = e_2 \cdot Ay \end{aligned}$$

where the last inequality follows by taking ε small enough. Next, we will show that any point in $\Gamma \setminus \{e_3\}$ is not an ESS. Recall that a point $\hat{x}$ is an Evolutionary Stable Strategy if for all $x \in \Delta \setminus \{\hat{x}\}$ one has for $\varepsilon > 0$ small enough that

$$x \cdot A(\varepsilon x + (1-\varepsilon)\hat{x}) < \hat{x} \cdot A(\varepsilon x + (1-\varepsilon)\hat{x}).$$

Fix $\varepsilon > 0$, let $\hat{x} = \begin{pmatrix} 1-p_3 \\ 0 \\ p_3 \end{pmatrix} = p$ where $p_3 \in [\frac{c}{b\omega}, 1)$, and let $x = e_3$ then

$$e_3 \cdot A(\varepsilon e_3 + (1-\varepsilon)p) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot A \begin{pmatrix} (1-\varepsilon)(1-p_3) \\ 0 \\ \varepsilon + (1-\varepsilon)p_3 \end{pmatrix} = -c + b\omega(\varepsilon + (1-\varepsilon)p_3)$$
$$p \cdot A(\varepsilon e_3 + (1-\varepsilon)p) = \begin{pmatrix} 1-p_3 \\ 0 \\ p_3 \end{pmatrix} \cdot A \begin{pmatrix} (1-\varepsilon)(1-p_3) \\ 0 \\ \varepsilon + (1-\varepsilon)p_3 \end{pmatrix} = -c + b\omega(\varepsilon + (1-\varepsilon)p_3).$$

So it follows that no point in $\Gamma \setminus \{e_3\}$ is an ESS. Similarly, we can show that e_3 is not an ESS either. Let $\begin{pmatrix} y_1 \\ 0 \\ 1-y_1 \end{pmatrix} = y \in \langle e_1, e_3 \rangle \setminus \{e_3\}$, then we have

$$y \cdot Ay = y_1(b\omega y_1 - c) + (1-y_1)(b\omega y_1 - c) = b\omega y_1 - c = e_3 \cdot Ay$$

which immediately tells us that e_3 is not an ESS either.

Finally, we will show that q is not an Evolutionary Stable Strategy. Recall that by Theorem 1.1 in the lecture notes any ESS is an asymptotically stable equilibrium for the replicator dynamics. This is not the case for q. Consider a point $\begin{pmatrix} 0 \\ x_2 \\ 1-x_2 \end{pmatrix} = x \in \langle e_2, e_3 \rangle$, then we have by the replicator equation

$$\begin{aligned} \dot{x}_2 &= x_2(0 - (1-x_2)(b\omega(1-x_2) - c + c\omega x_2)) \\ &= x_2(1-x_2)(\omega(b-c)x_2 - (c - b\omega)). \end{aligned}$$

Therefore, $\dot{x}_2$ is negative over $(0, \frac{b\omega - c}{(b-c)\omega})$, and positive over $(\frac{b\omega - c}{(b-c)\omega}, 1)$ which is equivalent to saying that q repels points on $\langle e_1, e_3 \rangle$. Since q is not asymptotically stable along $\langle e_1, e_3 \rangle$, then it cannot be an ESS.

We can conclude that the only Evolutionary Stable Strategy of this game is e_2.

Since we are still assuming $b\omega > c$, we can see in Figure 14 of the Lecture Notes that the simplex Δ is partitioned into two invariant subsets. This figure represents the replicator dynamics for $\frac{1}{1-\omega}\begin{pmatrix} b-c & -c & b-c \\ b & 0 & b(1-\omega) \\ b-c & -c(1-\omega) & b-c \end{pmatrix}$, which is obtained through adding or removing multiples of the vector $\mathbb{1}$ from A. As we have seen in Exercise 1.3 part 2., the replicator dynamics of these two matrices are identical. We will denote the two sets in the partition as

$$\begin{aligned} \Xi_d &:= \left\{ x \in \Delta \,|\, x_3 < \tfrac{c(1-\omega)}{(b-c)\omega} \right\} \\ \Xi_u &:= \left\{ x \in \Delta \,|\, x_3 > \tfrac{c(1-\omega)}{(b-c)\omega} \right\}. \end{aligned}$$

In order to simplify notation we will denote $\frac{c(1-\omega)}{(b-c)\omega}$ by $\hat{x}_3$. We will now quickly show that the line $\{x_3 = \hat{x}_3\}$ is invariant under the replicator dynamics

$$\begin{aligned}(Ax)_3 - x \cdot Ax_3 &= (Ax)_1 + c\omega x_2 - (Ax)_1(x_1 + x_3) - c\omega x_2 x_3 \\ &= (Ax)_1(1 - x_1 - x_3) + c\omega x_2(1 - x_3) \\ &= x_2((Ax)_1 + c\omega(1 - x_3)) \\ &= x_2(b\omega x_3 - c + c\omega - c\omega x_3) \\ &= x_2(x_3(\omega(b - c)) - c(1 - \omega)) \\ &= \frac{x_2}{\omega(b - c)}(x_3 - \hat{x}_3).\end{aligned}$$

Since $\dot{x}_3 = x_3((Ax)_3 - x \cdot Ax) = \frac{x_2 x_3}{\omega(b-c)}(x_3 - \hat{x}_3)$ then we see that on $\{x_3 = \hat{x_3}\}$ the derivative $\dot{x}_3$ is zero, which means that the flow is constrained along the line. Both Ξ_d and Ξ_u are invariant.

In Ξ_d we have only one Nash Equilibrium e_2, which is also an Evolutionary Stable Strategy. Therefore e_2 is an asymptotically stable equilibrium for the flow starting in Ξ_d. The point e_2 in the simplex corresponds to the strategy allD, hence the most optimal (and stable) strategy to play in Ξ_d is to always defect.

The situation is slightly more delicate in Ξ_u. Here we have a line of Nash Equilibria along $\langle e_1, e_3\rangle \cap \Xi_u$, namely Γ. Every point in Γ is an equilibrium for the replicator dynamics, and attracts point in the interior of Δ, as showed by the vector field in the left diagram in Figure 15 in the lecture notes. Therefore, depending on the Initial Value Problem the flow in Ξ_u can end up reaching a different point in Γ, meaning that the recommended strategy depends on the initial conditions one chooses, and it is a mixed strategy (between allC and TFT). The only time we can get a pure strategy is if one starts in $\langle e_2, e_3\rangle \cap \Xi_u$, then the flow tends to e_3, or TFT.

Exercise 3.4:
1) Consider the replicator dynamics defined by the matrix

$$A = \begin{pmatrix} 0 & -1 & \delta\sigma \\ 1 & 0 & -\kappa\sigma \\ \delta & -\kappa & 0 \end{pmatrix}$$

where $\delta = \omega\varepsilon, \kappa = 1 - \omega + \omega k\varepsilon, \sigma = \frac{b\theta - c}{c - c\theta}$, and $\theta = \omega(1 - (k+1)\varepsilon)$ are positive constants. The replicator equations associated with A are

$$\begin{aligned}\dot{x}_1 &= x_1(-x_2 + \delta\sigma x_3 - x_3(1+\sigma)(\delta x_1 - \kappa x_2)) \\ \dot{x}_2 &= x_2(x_1 - \kappa\sigma x_3 - x_3(1+\sigma)(\delta x_1 - \kappa x_2)) \\ \dot{x}_3 &= x_3(1 - x_3(1+\sigma))(\delta x_1 - \kappa x_2).\end{aligned}$$

Consider the Lyapunov function $P(x_1, x_2, x_3) = x_1^A x_2^B x_3^C(1 - (1+\sigma)x_3)$, where $A = \frac{\kappa}{\theta}, B = \frac{\delta}{\theta}$, and $C = -\frac{1}{\theta}$. We claim that P is constant along the orbits of the solution of the previous system of ODEs. In order to prove this we will show that the (logarithmic) derivative of P is zero. Hence,

$$\begin{aligned}
\frac{\dot{P}}{P}(x_1, x_2, x_3) &= \dot{\log P} = \frac{d}{dt}(A \log x_1 + B \log x_2 + C \log x_3 + \log(1 - (1+\sigma)x_3)) \\
&= A\frac{\dot{x}_1}{x_1} + B\frac{\dot{x}_2}{x_2} + C\frac{\dot{x}_3}{x_3} - \frac{(1+\sigma)\dot{x}_3}{1-(1+\sigma)x_3} \\
&= -[1 + A + B + C](x_3(1+\sigma)(\delta x_1 - \kappa x_2)) + \\
&\quad + [-Ax_2 + \delta\sigma A x_3 + Bx_1 - \kappa\sigma B x_3 + \delta C x_1 - \kappa C x_2] \\
&= 0
\end{aligned}$$

where we get zero in the last equality since

$$\begin{aligned}
1 + A + B + C =& 1 + \frac{\kappa}{\theta} + \frac{\delta}{\theta} - \frac{1}{\theta} \\
=& \frac{1 - \omega + \omega k\varepsilon + \omega\varepsilon - 1 + \omega - \omega k\varepsilon - \omega\varepsilon}{\theta} = 0
\end{aligned}$$

and

$$\begin{aligned}
-Ax_2 + \delta\sigma A x_3 + Bx_1 - &\kappa\sigma B x_3 + \delta C x_1 - \kappa C x_2 \\
&= -\frac{\kappa}{\theta}x_2 + \frac{\delta\sigma\kappa}{\theta}x_3 + \frac{\delta}{\theta}x_1 - \frac{\kappa\sigma\delta}{\theta}x_3 - \frac{\delta}{\theta}x_1 + \frac{\kappa}{\theta}x_2 = 0.
\end{aligned}$$

We can conclude that P is constant along the orbits of the system we wrote down at the beginning of this solution. Therefore, the level sets of the function P describe the shape of the flow in Δ.

4 The Best Response Dynamics

4.1 Rock-Scissor-Paper game and some other examples

Exercise 4.1:
a) Consider the matrix

$$A = \begin{pmatrix} 0 & 6 & -4 \\ -3 & 0 & 5 \\ -1 & 3 & 0 \end{pmatrix}$$

and the Lyapunov function $V(x) = \max_i (Ax)_i$. As illustrated in Example 4.2 in the Lecture Notes, the simplex Δ can be divided into three regions over which the Best Response is single-valued

$$\begin{aligned}
\Xi_1 &= \{x \in \Delta \mid \mathcal{BR}(x) = \{e_1\}\} \setminus (Z_{1,2} \cup Z_{1,3}) \\
\Xi_2 &= \{x \in \Delta \mid \mathcal{BR}(x) = \{e_2\}\} \setminus (Z_{1,2} \cup Z_{2,3}) \\
\Xi_3 &= \{x \in \Delta \mid \mathcal{BR}(x) = \{e_3\}\} \setminus (Z_{1,3} \cup Z_{2,3}).
\end{aligned}$$

Figure 22 in the Lecture Notes reports the level set $\{x \mid V(x) = 0\}$ which is given by the union of three segments (in light blue, in the right simplex of Figure 22). We will now show that those light blue segment are segments of lines in Δ. The blue segments are given by imposing $V(x) = 0$ in the three regions Ξ_1, Ξ_2, and Ξ_3. Indeed, in Ξ_1 we have that $V|_{\Xi_1}(x) = (Ax)_1 = 6x_2 - 4x_3$, hence the blue segment is given by

$$L_1 = \{6x_2 - 4x_3 = 0\} \cap \Xi_1,$$

where $\{6x_2 - 4x_3 = 0\}$ can be see as a plane in $\mathbb{R}^3$ intersecting the simplex Δ. Note that $e_1 \in \{6x_2 - 4x_3 = 0\}$. Following the same reasoning we get that in Ξ_2 the blue segment is given by

$$L_2 = \{-3x_1 + 5x_3 = 0\} \cap \Xi_2;$$

whereas in Ξ_3 it is given by

$$L_3 = \{-x_1 + 3x_2 = 0\} \cap \Xi_3.$$

Once again notice that $e_2 \in \{-3x_1 + 5x_3 = 0\}$, and $e_3 \in \{-x_1 + 3x_2 = 0\}$. We can see L_1, L_2, and L_3 as three lines in Δ going through e_1, e_2, and e_3 respectively, restricted to the appropriate regions where $\mathcal{BR}$ is single-valued.

b) The next step is to show that L_1, L_2, and L_3 are invariant under the flow in Ξ_1, Ξ_2, and Ξ_3, respectively. In order to see this we will show that the flow $\dot{x}$ restricted to L_i has the same direction as the line L_i. For example in Ξ_1 we have

$$\dot{x}|_{L_1} = \mathcal{BR}(x) - x|_{L_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 - \frac{5}{2}x_2 \\ x_2 \\ \frac{3}{2}x_2 \end{pmatrix} = \begin{pmatrix} \frac{5}{2}x_2 \\ -x_2 \\ -\frac{3}{2}x_2 \end{pmatrix}$$

and the normal vector to the plane $\{6x_2 - 4x_3 = 0\}$ is given by $\hat{n}_1 = \left(\begin{smallmatrix} 0 \\ 6 \\ -4 \end{smallmatrix}\right)$. By taking the dot product we see that

$$\dot{x}|_{L_1} \cdot \hat{n}_1 = -6x_2 + 4\frac{3}{2}x_2 = 0$$

which means that the flow $\dot{x}$ along L_1 has no normal component to L_1, therefore L_1 is invariant under the flow in Ξ_1. We can carry out similar calculations in Ξ_2, and Ξ_3. Indeed,

$$\dot{x}|_{L_2} = \mathcal{BR}(x) - x|_{L_2} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} x_1 \\ 1 - \frac{8}{5}x_1 \\ \frac{3}{5}x_1 \end{pmatrix} = \begin{pmatrix} -x_1 \\ \frac{8}{5}x_1 \\ -\frac{3}{5}x_1 \end{pmatrix}$$

and the normal to the plane $\{-3x_1 + 5x_3 = 0\}$ is given by

$$\hat{n}_2 = \begin{pmatrix} -3 \\ 0 \\ 5 \end{pmatrix}.$$

As before, the dot product between these two vectors is zero

$$\dot{x}|_{L_2} \cdot \hat{n}_2 = 3x_1 - 5\frac{3}{5}x_1 = 0.$$

Finally, in Ξ_3 we have

$$\dot{x}|_{L_3} = \mathcal{BR}(x) - x|_{L_3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 3x_2 \\ x_2 \\ 1 - 4x_2 \end{pmatrix} = \begin{pmatrix} -3x_2 \\ -x_2 \\ 4x_2 \end{pmatrix}$$

and the normal to the plane $\{-x_1 + 3x_2 = 0\}$ is given by

$$\hat{n}_3 = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix}.$$

As before, the dot product between these two vectors is zero

$$\dot{x}|_{L_3} \cdot \hat{n}_3 = 3x_2 - 3x_2 = 0.$$

We can conclude that the segment L_i in Ξ_i is invariant under the Best Response dynamics.

c) We will now turn our attention to the derivative of the function V. Let $x \in \Xi_i$ then we have that $\mathcal{BR}(x) = \{e_i\}$ hence

$$\dot{V}(x) = e_i \cdot A\dot{x} = e_i \cdot (\mathcal{BR}(x) - x) = e_i \cdot Ae_i - e_i \cdot Ax = A_{ii} - V(x) = -V(x)$$

since $A_{ii} = 0$. Hence $\dot{V} = -V$ in the regions where the Best Response is single-valued and constant.

4)The ODE we have just computed $\dot{V}(x) = -V(x)$, only holds for $x \in \bigcup_i \Xi$. This equation, together with IVP $V(x(0)) = V(x_0)$ for $x_0 \in \Xi_i$ tells that $V(x) = V(x_0)e^{-t}$, as long as $x \in \bigcup_i \Xi_i$, which seems to suggest that the solution to this ODE tends to zero exponentially fast. Unfortunately the solutions will have to cross the indifference, and along such lines the equality $\dot{V} = -V$ will not hold. For example, along $Z_{2,3}$ we have that $\dot{V}(x) = \frac{d}{dt}\max_i Ax_i = e_2 \cdot A\dot{x} = e_2 \cdot A(\mathcal{BR}(x) - x)$ since $(Ax)_2 = (Ax_3)$ and they are maximal. We immediately see that now $\dot{V}$ is multi-valued, and therefore we cannot have $\dot{V} = -V$.

In conclusion V decays to 0 exponentially fast, as long as it does not hit an indifference line. Unfortunately, V tends to 0 as t tends to infinity, and a crossing of an indifference line is unavoidable.

4.2 Two player best response dynamics

Exercise 4.2:
We are now interested in the Best Response dynamics for two players. We consider the matrices

$$A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the system of (possibly multivalued) ODEs

$$\begin{aligned} \dot{x} &= \mathcal{BR}_A(y) - x \\ \dot{y} &= \mathcal{BR}_B(x) - y. \end{aligned}$$

In Example 4.3 we proved that this system has a unique NE in the interior of Δ, namely $E = (E^A, E^B) = \left(\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\right)$, and that the Lyapunov function $V(x, y) = \mathcal{BR}_A(y) \cdot Ay + x \cdot B\mathcal{BR}_B(x)$ is such that $V(x, y) \geq V(E^A, E^B) = 0$. Let us adopt the second convention. By a simple computation we can see that

$$\mathcal{BR}_A(y) = \begin{cases} \{e_2\} & \text{if } y_1 > y_2 \\ \Delta & \text{if } y_1 = y_2 = \frac{1}{2} \\ \{e_1\} & \text{if } y_1 < y_2 \end{cases} \qquad \mathcal{BR}_B(x) = \begin{cases} \{e_1\} & \text{if } x_1 > x_2 \\ \Delta & \text{if } x_1 = x_2 = \frac{1}{2} \\ \{e_2\} & \text{if } x_1 < x_2 \end{cases}$$

which immediately tells us that E is the only NE in general. Notice that $V(x, y) = 0$ if and only if (x, y) is a NE for (A, B). We know that if we supplement the ODE $\dot{V} = -V$ with the initial condition $(x(0), y(0)) = (x_0, y_0)$, we get the solution $V(x(t), y(t)) = e^{-t}V(x_0, y_0)$. Assuming that our initial condition is not E, then $V(x_0, y_0) > 0$, which means that $V(x(t), y(t))$ reaches 0 as $t \to \infty$.

In Example 4.1, we find ourselves in a similar situation as we just described. Again, we have that $V(x) > V(E)$ for $x \in \Delta \setminus \{E\}$, and that $\dot{V} = -V$, hence $V(x(t)) = e^{-t}V(x_0)$. The main difference is that now $V(E) = \frac{a-b}{3}$, so if we assume that $a > b$ then $V(E) > 0$. This means that if we assume $x_0 \neq E$, our solution starting at x_0 reaches E in finite time $t = \ln\left(\frac{V(x_0)}{V(E)}\right) < \infty$. If $a = b$

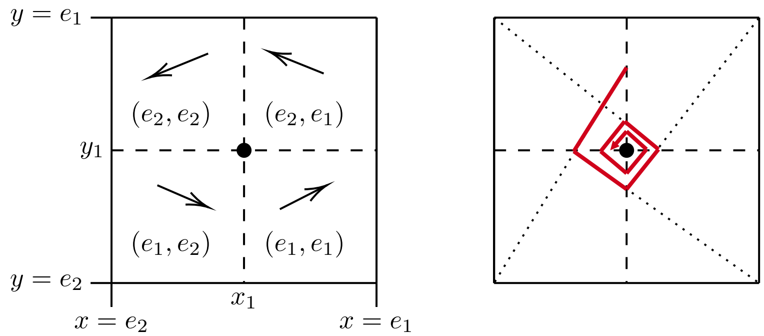


Figure 22: In the left square we reported the direction of the flow, and the values of the Best Response map in the regions where it is constant. Notice we plotted x_1 on the horizontal axis, against y_1 on the vertical one. In the right square we plotted the flow converging to the equilibrium $(\frac{1}{2}, \frac{1}{2})$.

then the solution converges to E as time tends to infinity. If $a < b$ we have the appearance of the Shapley triangle to which our solution converges to, as explained in Example 4.1.

As we have just seen, the flow associated to the Best Response dynamic for the game described by A and B tends towards to $((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2})$ as showed in Figure 22. The velocity of the flow does not go to zero! We will denote the four regions in the squares in Figure 22 using cardinal directions: starting from the top right region, and moving anticlockwise we have North-West (NW), South-West (SW), South-East (SE), and North-East (NE). Consider the NW region $[0, \frac{1}{2}] \times [\frac{1}{2}, 1]$, here the Best Response function is single valued and equals (e_2, e_2) which leads to the system

$$\dot{x} = \begin{pmatrix} -x_1 \\ 1 - x_2 \end{pmatrix} = \begin{pmatrix} -x_1 \\ x_1 \end{pmatrix}$$
$$\dot{y} = \begin{pmatrix} -y_1 \\ 1 - y_2 \end{pmatrix} = \begin{pmatrix} -y_1 \\ y_1 \end{pmatrix}.$$

Hence the velocity v_{NW} in this quadrant is given by

$$|v_{\text{NW}}|^2 = |\dot{x}|^2 + |\dot{y}|^2 = 2x_1^2 + 2y_1^2$$

which tends to 1 as the flow tends to its equilibrium. This velocity is always strictly positive in this quadrant.

Similar we obtain the velocities

$$v_{\text{SW}} = 2x_2^2 + 2y_1^2$$
$$v_{\text{SE}} = 2x_2^2 + 2y_2^2$$
$$v_{\text{NE}} = 2x_1^2 + 2y_2^2$$

which are always strictly greater than zero if our flow starts in the interior of the phases space, and all tend to 1 as the solution tends to its equilibrium.

4.3 Convergence and non-convergence to Nash Equilibrium for Best Response Dynamics

Exercise 4.3:
1) Let A be 3×3 matrix, let $\alpha, \beta, \gamma \in \mathbb{R}$, and $c > 0$. We will show that for any $y \in \Delta$, then $\mathcal{BR}_{A'}(y) = \mathcal{BR}_A(y)$ where

$$A' \coloneqq cA + \begin{pmatrix} \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \end{pmatrix}.$$

This statement can be proved through a direct calculation

$$\begin{aligned}
\mathcal{BR}_{A'}(y) &= \arg\max_{x\in\Delta} x \cdot A'y = \arg\max_{x\in\Delta} x \cdot \left(cA + \begin{pmatrix} \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \end{pmatrix} \right) y \\
&= \arg\max_{x\in\Delta} \left[x \cdot (cA)y + x \cdot \begin{pmatrix} \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \\ \alpha & \beta & \gamma \end{pmatrix} y \right] \\
&= \arg\max[c(x \cdot Ay) + (\alpha y_1 + \beta y_2 + \gamma y_3)(x_1 + x_2 + x_3)] \\
&= c \arg\max_{x\in\Delta}(x \cdot Ay) + \alpha y_1 + \beta y_2 + \gamma y_3 \\
&= \mathcal{BR}_A(y),
\end{aligned}$$

as we claimed.

2) Consider the two matrices

$$A = \begin{pmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{pmatrix} \qquad B = \begin{pmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{pmatrix}$$

we will show that we for $\beta = \phi = \frac{\sqrt{5}-1}{2}$, the reciprocal of the golden ratio, the matrix B can be rescaled to give a zero-sum game. A quick remark on the chosen value for β

$$\phi^2 = \frac{6 - 2\sqrt{5}}{4} = 1 + \frac{1 - \sqrt{5}}{2} = 1 - \phi. \tag{9}$$

Consider the matrix

$$\tilde{B} = \phi \left(B - \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \right) = \begin{pmatrix} -\phi(\beta+1) & 0 & -\phi \\ -\phi & -\phi(\beta+1) & 0 \\ 0 & -\phi & -\phi(\beta+1) \end{pmatrix}$$

which by the first part of this question we know yields the same Best Response (and therefore Best Response dynamics) as B. Clearly we have

$$A + \tilde{B} = \begin{pmatrix} 1-\phi(\beta+1) & 0 & \beta - \phi \\ \beta - \phi & 1 - \phi(\beta+1) & 0 \\ 0 & \beta - \phi & 1 - \phi(\beta+1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

since $\beta = \phi$, and $1 - \phi(1+\beta) = 1 - \phi - \phi^2 = 0$ by Equation 9.

3) This solution is adapted from the lecture notes. Take the Shapley periodic orbit $\gamma : \mathbb{R} \to \Delta \times \Delta \subset \mathbb{R}^6$ of the Best Response dynamics associated to the two player Rock–Paper–Scissors game corresponding to $\beta = 0$. Note that $\mathcal{BR}_A(e_i) = e_i$ and $\mathcal{BR}_B(e_i) = e_{i+1}$ (note that we using the 2^{nd} notation for the matrices). Let π_A, π_B be the projections of $\Delta_A \times \Delta_B \subset \mathbb{R}^6$ onto the two triangles shown in Figure 23. The blue triangles drawn in this figure correspond

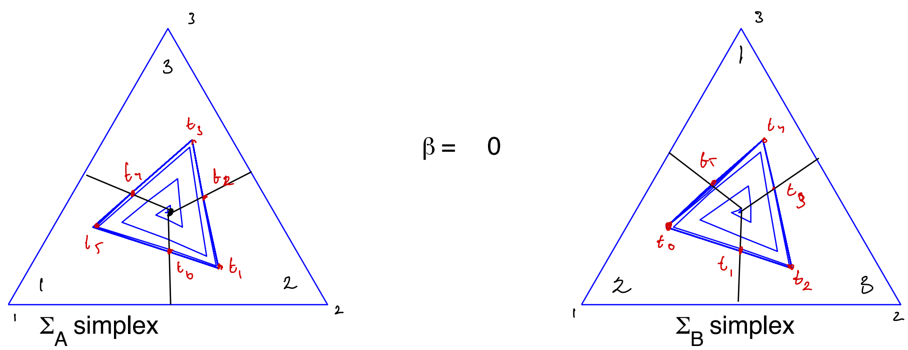


Figure 23: The projection of the Shapley's periodic orbit γ onto the two simplices Δ_A and Δ_B. The numbers inside the triangle Δ_A denote to corner to which player B will be heading, and the number inside the triangle Δ_B where player A is heading.

to the projections $\pi_A(\gamma)$ and $\pi_B(\gamma)$ of γ. Let T be the period of γ and let $0 = t_0 < t_1 < \cdots < t_5 < t_6 = T$ be the times when $\pi_A(\gamma(t))$ or $\pi_B(\gamma(t))$ are contained in one of the indifference lines. When $\pi_A(\gamma(t))$ lies in an indifference line at $t = t'$, then $t \mapsto \pi_B(\gamma(t))$ changes from moving towards one corner for $t < t'$ close to t' to moving towards another corner for $t > t'$ close to t'. In fact, for each t we have that $\gamma(t)$ intersects at most one of the indifference lines. The points $\pi_A(\gamma(t_i))$ and $\pi_B(\gamma(t_i))$ are indicated in the figure, and note that the points move anti-clockwise in the triangles and head towards e_i in Δ_A when $\gamma(t)$ is in the region in Δ_B marked with i (and vice versa). The curve γ is a solution of the piece-wise smooth ODE:

$$\dot{\gamma}(t) = \begin{pmatrix} e_i \\ e_j \end{pmatrix} - \gamma(t) \text{ for } t \in (t_i, t_{i+1}).$$

Here $\left(\begin{smallmatrix} e_i \\ e_j \end{smallmatrix}\right)$ are best response choices. i.e., $e_i = \mathcal{BR}_A(\pi_B(\gamma(t))$ and $e_j = \mathcal{BR}_B(\pi_A(\gamma(t))$. The solution of this ODE is

$$\gamma(t) = (1 - e^{-t}) \begin{pmatrix} e_i \\ e_j \end{pmatrix} + e^{-t}\gamma(0) \text{ for } t \in (t_i, t_{i+1}). \qquad (10)$$

So $(t_i, t_{i+1}) \ni t \mapsto \gamma(t)$ is a straight line in $\mathbb{R}^6$ (which is contained in $\Delta_A \times \Delta_B$). Let us take a closer look at what these lines: the best response choices depend on time in the following way

$$\begin{pmatrix} e_i \\ e_j \end{pmatrix} = \begin{cases} (e_2, e_2)^\mathsf{T} & \text{if } t \in (t_0, t_1) \\ (e_3, e_2)^\mathsf{T} & \text{if } t \in (t_1, t_2) \\ (e_3, e_3)^\mathsf{T} & \text{if } t \in (t_2, t_3) \\ (e_1, e_3)^\mathsf{T} & \text{if } t \in (t_3, t_4) \\ (e_1, e_1)^\mathsf{T} & \text{if } t \in (t_4, t_5) \\ (e_2, e_1)^\mathsf{T} & \text{if } t \in (t_5, t_0). \end{cases}$$

We will now see that all these vectors describe a regular hexagon. First of all we know that γ is composed of six sides and that the length of the first two sides is given by

$$\gamma(t_1) - \gamma(t_0) = \lambda((e_2, e_2) - \gamma(t_0))$$

and

$$\gamma(t_2) - \gamma(t_1) = \lambda((e_3, e_2) - \gamma(t_1)),$$

where λ is as in the notes. Hence, using the formulas for $\gamma(t_i)$

$$||\gamma(t_1) - \gamma(t_0)|| = \lambda||(e_2, e_2) - \gamma(t_0)|| = \lambda C||(\theta^3, \theta^3 - C, \theta, \theta^4, 1 - C, \theta^2)||$$

and

$$||\gamma(t_2) - \gamma(t_1)|| = \lambda||(e_3, e_2) - \gamma(t_1)|| = \lambda C||(\theta^2, \theta^4, 1 - C, \theta^3, \theta^3 - C, \theta)||$$

and this implies that

$$||\gamma(t_1) - \gamma(t_0)|| = ||\gamma(t_2) - \gamma(t_1)||$$

and therefore by symmetry

$$||\gamma(t_{i+1}) - \gamma(t_i)|| = ||\gamma(t_1) - \gamma(t_0)||$$

for all i.

5 Fictitious Play: a Learning Model

5.1 Best response and fictitious play

Exercise 5.1:
1) Consider the fictitious play dynamics

$$\begin{aligned}\dot{p}(s) &= \frac{1}{s}(\mathcal{BR}_A(q(s)) - p(s))\\ \dot{q}(s) &= \frac{1}{s}(\mathcal{BR}_B(p(s)) - q(s))\end{aligned}$$

where the quantities p, and q depend on the previously played strategies, as explained in the lecture notes. This is a non-autonomous systems of ODEs (notice the factor $\frac{1}{s}$ in the RHS!), but this can be reduced to an autonomous system if we substitute $s = e^t$. Indeed, if we define $\tilde{p}(t) = p(e^t)$, and $\tilde{q}(t) = q(e^t)$ then by imposing $s = e^t$ we get the following system of ODEs

$$\begin{aligned}\dot{\tilde{p}}(t) &= e^t\dot{p}(e^t) = e^t\frac{1}{e^t}(\mathcal{BR}_A(q(e^t)) - p(e^t)) = \mathcal{BR}_A(\tilde{q}(t)) - \tilde{p}(t)\\ \dot{\tilde{q}}(t) &= e^t\dot{q}(e^t) = e^t\frac{1}{e^t}(\mathcal{BR}_B(p(e^t)) - q(e^t)) = \mathcal{BR}_B(\tilde{p}(t)) - \tilde{q}(t)\end{aligned}$$

which describes the Best Response dynamics. If we consider the Shapley dynamics of Example 4.5 then we have that the orbits of the fictitious play $(p(t), q(t))$ are the same as the orbits of the best response $(\tilde{p}(t), \tilde{q}(t))$, since we have only smoothly reparameterised time.

2) For $\beta = 0$ we know that the Shapley Best Response dynamics shows the presence of closed periodic orbits, i.e.

$$(\tilde{p}(t), \tilde{q}(t)) = (\tilde{p}(t+T), \tilde{q}(t+T)) \tag{11}$$

where T represents the period, or the amount of time needed to complete a full lap of the periodic orbit. Since the fictitious play can be seen as a reparameterisation of the Best Response dynamics then the system present the same closed periodic orbit, but its characterisation is slightly different. The speed along such orbit decreases, which means that it will take us longer, and longer to complete a full lap of the orbit. Indeed, instead of having a characterisation like in Equation 11, we have

$$\begin{aligned}(\tilde{p}(t), \tilde{q}(t)) &= (\tilde{p}(t+T), \tilde{q}(t+T))\\ \iff (p(e^t), q(e^t)) &= (p(e^{t+T}), q(e^{t+T}))\\ \iff (p(s), q(s)) &= (p(e^T s), q(e^T s))\end{aligned}$$

which exactly tells us that it will take us exponentially longer to complete a full lap compared to the precedent lap.

5.2 The no–regret set

Exercise 5.2:
Try to understand how to the CCE set from Chapter 5 and the CE set from Chapter 7 are related. Why do we say that CCE $\subseteq$ CE? What probability distribution would you be more inclined to follow and why?

5.3 Fictitious play converges to the no–regret set CCE

Exercise 5.3:
Open ended question.

5.4 FP orbits often give better payoff than Nash

Exercise 5.4:
1) Consider the two matrices

$$A = A_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad B = B_0 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

giving rise to the Shapley system for the Best Response dynamics. As we have seen in Exercise 5.1, the periodic orbit (or Shapley triangles) under the $\mathcal{BR}$ dynamics that was analysed in Example 4.5 in the lecture notes corresponds to a closed orbit γ for the FP dynamics (note that γ is periodic under the BR dynamics). By the very own definition of closed orbit, if the initial condition for our FP dynamics is along such an orbit, then the flow is constrained there. Therefore the limits of the points $\gamma(t)$ lie in Shapley's periodic orbit. By Theorem 5.1, the probability distribution

$$p_{ij}(t) = \frac{1}{t}\int_0^t x_i(s)y_j(s)ds$$

converges to the CCE set.

From previous analyses, we know that the point $(\hat{p}, \hat{q}) = \left(\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)\right)$ is a Nash Equilibrium for the system. By Lemma 5.1 we know that the every element in the set of Nash Equilibria corresponds an element in the CCE set. Therefore, the CCE set for the Shapley system with $\beta = 0$ is composed of at least two elements.

The last thing we are left to check is that, indeed, these two probability matrices are different. The element in the CCE set given by the NE equilibrium $(\hat{p}, \hat{q})$ is simply

$$\hat{P} = \begin{pmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9}. \end{pmatrix}$$

We will now show that some of the entries of the probability matrix $P(t)$ associated to the Shapley orbit of the FP dynamics have value different from $\frac{1}{9}$.

We will use the same notation and some of the results in part 3 of Exercise 4.3. Let $\gamma(t) = (p(t), q(t))$ be the Shapley periodic orbit, and suppose that to complete a lap of this periodic orbit takes e^T time, i.e. $\gamma(t_0) = (p(t_0), q(t_0)) = (p(e^T t_0), q(e^T t_0)) = \gamma(e^T t_0)$.

One approach to this is to analyse the strategies played in terms of the FP dynamics. Recall that we the strategies played under the FP dynamics are given by

$$x(t) \in \mathcal{BR}_A(q(t))$$
$$y(t) \in \mathcal{BR}_B(p(t)$$

and notice that since we are restricting our attention to a flow over the closed orbit γ then we have that the Best Response function is piece-wise constant outside of 6 points on γ (the corners of the hexagon) and it is equal to

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{cases} (e_2, e_2)^\mathsf{T} & \text{if } t \in (t_0, t_1) \\ (e_3, e_2)^\mathsf{T} & \text{if } t \in (t_1, t_2) \\ (e_3, e_3)^\mathsf{T} & \text{if } t \in (t_2, t_3) \\ (e_1, e_3)^\mathsf{T} & \text{if } t \in (t_3, t_4) \\ (e_1, e_1)^\mathsf{T} & \text{if } t \in (t_4, t_5) \\ (e_2, e_1)^\mathsf{T} & \text{if } t \in (t_5, t_0) \end{cases}.$$

We are now seeing the orbit γ as a closed periodic orbit of period T. If we now want to compute the entry $p_{1,3}$ of the CCE distribution P given by γ we have

$$p_{1,2}(T) = \frac{1}{T} \int_0^T x_1(t) y_2(t)\, dt = 0$$

since there exists no time where both the 1^{st} entry of $x(t)$ and the 2^{nd} entry of $y(t)$ are simultaneously non-zero. This comes from the very specific form of the Best Response function along the periodic orbit: if $x(t) = e_i$ then $y(t)$ can only be equal to e_i or e_{i-1}, and never e_{i+1} (the indexes are to be take $\mod 3$).

Another approach is to note that from

$$p_{ij}(t) = \frac{1}{t} \int_0^t x_i(s) y_j(s) ds$$

it follows that $\sum_j p_{ij}(t) = \frac{1}{t} \int_0^t x_i(s) = p(t)$ and $\sum_i p_{ij}(t) = \frac{1}{t} \int_0^t y_j(s) = q(t)$ where $p(t)$ and $q(t)$ are so that $\gamma(t) = (p(t), q(t))$. It follows that the marginals of the probability matrix $P(t)$ do not converge as $t \to \infty$.

2) Suppose that our solution for the FP dynamics associated to the Shapley system is bound to the periodic orbit we precedently discussed. By Proposition 5.1 we know that

$$\lim_{t \to \infty} \hat{u}^A(T) - \max_{\overline{p} \in \Delta} \overline{p} \cdot Aq(T) = 0$$
$$\lim_{t \to \infty} \hat{u}^B(T) - \max_{\overline{q} \in \Delta} p(T) \cdot B\overline{q} = 0.$$

The orbit $(p(s), q(s))$ is constrained along the Shapley triangle, and from Exercise 5.1 we know that $(p(s), q(s)) = (p(e^K s), q(e^K s))$, for $K > 0$. This means that the quantities $\max_{\overline{p}\in\Delta} \overline{p} \cdot Aq(T)$ and $\max_{\overline{q}\in\Delta} p(T) \cdot B\overline{q}$ just need to be computed over one lap of the flow along the Shapley periodic orbit instead for all times. Therefore

$$\lim_{T\to\infty} \max_{\overline{p}\in\Delta} \overline{p} \cdot Aq(T) = \max_{T\in[t_0, e^K t_0]} \max_{\overline{p}\in\Delta} \overline{p} \cdot Aq(T) = c_A$$
$$\lim_{T\to\infty} \max_{\overline{q}\in\Delta} p(T) \cdot B\overline{q} = \max_{T\in[t_0, e^K t_0]} \max_{\overline{q}\in\Delta} p(T) \cdot B\overline{q} = c_B$$

by compactness. We can conclude that the average payoffs for FP dynamics flows along the Shapley periodic orbit converge

$$\lim_{t\to\infty} \hat{u}^A = c_A \qquad \lim_{t\to\infty} \hat{u}^B = c_B.$$

5.5 Discrete fictitious dynamics

Exercise 5.5:
Try to adapt the algorithms in Appendix B for the discrete FP dynamics of the Shapley system.

6 Reinforcement Learning

6.1 Set-up of reinforcement learning

Exercise 6.1:
1) As explained in the exercise let a doctor be prescribing either a Placebo or a Medicine to patients of type I or II. Suppose that the matrix describing this scenario is given by

$$\begin{array}{cc} & \begin{array}{cc} \mathrm{I} & \mathrm{II} \end{array} \\ \begin{array}{c} \mathrm{M} \\ \mathrm{P} \end{array} & \begin{pmatrix} 10 & 0 \\ 5 & 5 \end{pmatrix} \end{array}.$$

Let q be the probability of a patient being of type I and $1-q$ of type II. Then assuming that the doctor prescribes M with probability p and P with probability $1-p$ the payoff is

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} \cdot \begin{pmatrix} 10 & 0 \\ 5 & 5 \end{pmatrix} \begin{pmatrix} q \\ 1-q \end{pmatrix} = \begin{pmatrix} p \\ 1-p \end{pmatrix} \cdot \begin{pmatrix} 10q \\ 5 \end{pmatrix}$$

So if $q < 1/2$ then taking $p = 0$ (i.e. prescribing a placebo) gives the best payoff whereas if $q > 1/2$ then taking $p = 1$ (i.e. prescribing the medicine) gives the best payoff.

2) The Python code and graphs indicating what outcomes to expect when the doctor uses the Erev-Roth model to determine which medication to prescribe can be found in the appendix of the lecture notes.

6.2 The Arthur model in the 2×2 setting

Exercise 6.2:
1) Let us start with the coordination game first. We wish to show that the set of singularities $S = \{(0,0),(0,1),(1,0),(1,1),(\theta_1,\theta_2)\} \subset [0,1]\times[0,1]$ is internally chain recurrent. The dynamics in this case is given by the system of replicator equations

$$\begin{aligned} \dot{x} &= x(1-x)[\alpha_1 - y(\alpha_1+\alpha_2)] \\ \dot{y} &= y(1-y)[\beta_1 - x(\beta_1+\beta_2)] \end{aligned}$$

and the point $\theta = (\theta_1,\theta_2) = \left(\frac{\beta_1}{\beta_1+\beta_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}\right)$ is an internal Nash Equilibrium. Let us denote by $\phi_t = (x(t), y(t))$ the flow for this system. Notice that for any point $s \in S$ we have that $\dot{x}|_s = \dot{y}|_s = 0$. This means that the set S is composed of fixed points, i.e. $\phi_t(s) = s$ for all $t \geq 0$. Notice this makes S automatically invariant.

We will now show that if s is a fixed point, then s is chain recurrent. This means that for all $\delta, T > 0$ there exists a (δ, T)-pseudo orbit of ϕ_t connecting s to itself. Fix $\delta, T > 0$ then let $x_0 = s$ and $t_0 = T+1$ then we simply have that $\phi_{t_0}(x_0) = \phi_{T+1}(s) = s$, hence

$$t_0 > T \qquad \text{and} \qquad d(\phi_{t_0}(x_0), x_0) = 0 < \delta,$$

as we wanted. Thus, every point in S is chain recurrent, making S internally chain recurrent.

Let us now move to the second case. We will now generalise what we have just showed. In the classification of 2×2 games this second system we are considering is a zero sum case with an internal Nash Equilibrium. In this case we are asked to show that the entire phase space $M = [0,1] \times [0,1]$ is internally chain recurrent. The main observation that has to be made here is that the phase space is foliated by periodic orbits of ϕ_t: for any $x \in M$ there exists a time $t(x)$, called period, such that $\phi_{t(x)}(x) = x$. We will denote the periodic orbit passing through x as $\gamma_x = \phi_{[0,t(x)]}(x)$. If $x \in \partial([0,1] \times [0,1])$ then $\gamma_x = \partial([0,1] \times [0,1])$. Notice that the existence of these periodic orbits is given by the last part of section 3 of Exercise 2.4.

We will now show that any point belonging to a periodic orbit is chain recurrent. Let $x \in M$, and fix $\delta, T > 0$. We will now define $x_0 = x$ and if $t(x) < T$ then we will let $t_0 = T + (t(x) - k)$ where $k = T \mod t(x)$ (clearly $k, T, t(x) \in \mathbb{R}_{\geq 0}$), otherwise $t_0 = 2t(x)$: in both cases $t_0 > T$, and $t(x)$ divides t_0. Therefore we have that

$$t_0 \geq T \qquad \text{and } d(\phi_{t_0}(x_0, x_0) = d(\phi_{lt(x)}(x), x) = d(x,x) = 0 < \delta$$

where $l = \frac{T}{t(x)} \in \mathbb{Z}$. We can conclude that every point in M is chain recurrent, and therefore M is internally chain recurrent.

2) Recall the assumption of Proposition 6.1: $a_{ij}, b_{i1j>C>0}$ for all i, j. First of all we want to show that there exists $\alpha, \alpha' > 1$ such that

$$1 - \frac{\alpha}{t} < 1 - \frac{a_{1j}}{Ct + a_{1j}} < 1 - \frac{\alpha'}{t}$$

for t big enough. We can rewrite the two inequalities as

$$\frac{\alpha'}{t} < \frac{a_{1j}}{Ct + a_{1j}} < \frac{\alpha}{t}.$$

We will prove them in order, left to right.

By assumption we know that $a_{ij} > C$ for all i, j, hence define $\varepsilon := \min_{i,j} a_{ij} - C > 0$. For any $\delta \in (0, \varepsilon)$ notice the following

$$\frac{a_{1j}}{C + \delta} > \frac{a_{1j}}{C + \varepsilon} = \frac{a_{1j}}{\min_{i,j} a_{i,j}} \geq 1,$$

for any j. Now fix a positive δ smaller than ε, then there exists a time t_0 such that $\frac{\max_{i,j} a_{ij}}{t} \leq \delta$ for all $t \geq t_0$. Therefore

$$\frac{a_{1j}}{Ct + a_{1j}} = \left(\frac{a_{1j}}{C + a_{1j}/t}\right)\frac{1}{t} \geq \left(\frac{a_{1j}}{C + \max_{i,j} a_{i,j}/t}\right)\frac{1}{t} \geq \left(\frac{a_{1j}}{C + \delta}\right)\frac{1}{t} \geq \left(\frac{\min_{i,j} a_{ij}}{C + \delta}\right)\frac{1}{t}$$

for $t \geq t_0$, and if we define $\alpha' := \frac{\min_{i,j} a_{i,j}}{C+\delta} > 1$ we can conclude that

$$\frac{a_{1j}}{Ct + a_{1j}} \geq \frac{\alpha'}{t}.$$

Notice that we could tweak the denominator in the definition of α' in order to get a sharp inequality above.

For the second inequality it is enough to remember that $a_{1j} > 0$, and that $\frac{a_{ij}}{C} > 1$ for all i, j. Indeed if we let $\alpha = \frac{\max_{i,j} a_{ij}}{C} > 1$ we obtain

$$\frac{a_{1j}}{Ct + a_{1j}} < \frac{a_{1j}}{Ct} \leq \left(\frac{\max_{i,j} a_{ij}}{C}\right)\frac{1}{t} = \frac{\alpha}{t}.$$

We can therefore conclude that for $t \geq t_0$

$$1 - \frac{\alpha}{t} < 1 - \frac{a_{1j}}{Ct + a_{1j}} < \frac{\alpha'}{t},$$

as we wanted. Recall that we define the sequence $(d^t)_t$ as $d^{t+1} = \left(\frac{a_{1j}}{Ct+a_{1j}}\right) d^t$, we will now mimic this construction. Define the sequence $(\tilde{d}^t)_t$ as

$$\begin{aligned} \tilde{d}^1 &:= d^1 \\ \tilde{d}^{t+1} &:= \left(1 - \frac{\alpha'}{t}\right)\tilde{d}^t \qquad \text{for all } t. \end{aligned}$$

Clearly

$$\frac{d^{t+1}}{d^t} = 1 - \frac{a_{1j}}{Ct + a_{1j}} < 1 - \frac{\alpha'}{t} = \frac{\tilde{d}^{t+1}}{\tilde{d}^t} \tag{12}$$

for all $t \geq t_0$. Similarly we can compare the two sequences: if $d^{t_0} \leq \tilde{d}^{t_0}$ then we automatically have that $d^t \leq \tilde{d}^t$ for all $t \geq t_0$, if, on the other hand, $d^{t_0} > \tilde{d}^{t_0}$ then there exists a time $\tau \geq t_0$ for which $d^t \leq \tilde{d}^t$ for all $t \geq \tau$ thanks to Inequality 12. Therefore, if we can show that the series $\sum_{t=1}^{\infty} \tilde{d}^t$ converges, then this will give us control over the two series tails $\sum_{t=t_0}^{\infty} d^t$ and $\sum_{t=\tau}^{\infty} d^t$ which will tell us that $\sum_{t=1}^{\infty} d^t$ converges.

In order to show that $\sum_{t=1}^{\infty} d^t$ converges we will use the Raabe test. In order for such test to work we need to check its two conditions.

1. $\lim_{t\to\infty} \left|\frac{\tilde{d}^t}{\tilde{d}^{t+1}}\right| = \lim_{t\to\infty} \frac{1}{1-\frac{\alpha'}{t}} = 1$, as we need;

2. $\lim_{t\to\infty} t\left(\left|\frac{\tilde{d}^t}{\tilde{d}^{t+1}}\right|\right) = \lim_{t\to\infty} \frac{\alpha' t}{t-\alpha'} = \alpha' > 1$;

since the second limit tends to a finite value bigger than 1 we can conclude that by the Raabe test the series $\sum_{t=1}^{\infty} \tilde{d}^t$ converges. Therefore, the series $\sum_{t=1}^{\infty} d^t$ converges, and this concludes our proof.

6.3 The Erev-Roth model

Exercise 6.3:
1) Consider the system of differential equations

$$\begin{aligned} \dot{p}_i &= \frac{p_i}{a(t)}[(Aq)_i - p \cdot Aq] \\ \dot{q}_j &= \frac{q_j}{b(t)}[(Bq)_j - q \cdot Bp] \\ \dot{a} &= -a + p \cdot Aq \\ \dot{b} &= -b + q \cdot Bp \end{aligned}$$

and we wish to study the singularities of it. Assume that p and q are n dimensional probability vectors. Firstly notice that in order to have a singularity we need

$$\begin{aligned} a &= p \cdot Aq \\ b &= q \cdot Bp \end{aligned}$$

where a, b are constant (since $\dot{a} = \dot{b} = 0$ for all times). In order to avoid complicated behaviours which will requite a much more in depth analysis we will assume that a and b are nonzero. The other $2n$ ODEs are always zero whenever

$$\begin{aligned} \dot{p}_i = 0 &\iff \begin{cases} p_i = 0 \\ (Ap)_i = p \cdot Aq & \text{if } p_i \neq 0 \end{cases} \\ \dot{q}_j = 0 &\iff \begin{cases} q_j = 0 \\ (Aq)_j = q \cdot Bp & \text{if } q_j \neq 0 \end{cases} \end{aligned}$$

for all $i, j = 1, 2, \ldots, n$. Since p and q are probability vectors, there exists at least one entry p_i and one entry q_j which are nonzero, meaning that at least for those indexes $(Ap)_i = p \cdot Aq = a$ and $(Bq)_j = q \cdot Bp = b$.

To conclude the singularities of the aforementioned system of ODEs are given by the simultaneous system of equations

$$\begin{aligned} (Ap)_i &= p \cdot Aq \\ (Bq)_j &= q \cdot Bp \\ a &= p \cdot Aq \\ b &= q \cdot Bp. \end{aligned}$$

for all i such that $p_i \neq 0$ and for all j such that $q_j \neq 0$. Unfortunately we are not able to produce anything more insightful about the location of the singularities for such a general system.

2) As we have seen before, approaching this problem from a purely theoretical

point of view will not take us far. A computational approach to this type of problems is usually preferred. Try to play around with different matrices and see how different singularities can arise in different places. If you use 3×3 matrices you can use the visualisation code you developed in Exercise 2.5 to understand where the singularities lie in the phase space Δ.

6.6 Q-Learning with softmax

Exercise 6.4:
For some more background on how these picture have been obtained we wish to redirect you to the paper *Frequency Adjusted Multi-agent Q-learning* by Michael Kaisers and Karl Tuyls (In Proc. of 9th Intl. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010), pp.309–315).

7 No Regret Learning

7.1 The correlated equilibrium (CE) set

Exercise 7.1:
1) Let (p,q) be a Nash Equilibrium for a game determined by the matrices (A,B). If we define our probability distribution matrix $S=(s_{ij})=p_iq_j$, then in order to check if this is a Correlated Equilibrium it is just a matter of working through the definition of CE, that we will now rewrite in vector form. In order to check if S is a CE we would need to show

$$\sum_k a_{i'k}s_{ik} = \sum_k a_{i'k}p_iq_k = p_i(Aq)_{i'} \le p_i(Aq)_i = \sum_k a_{ik}p_iq_k = \sum_k a_{ik}s_{ik}$$
$$\sum_k b_{kj'}s_{kj} = \sum_k b_{kj'}p_kq_j = q_j(p^\mathsf{T}B)_{j'} \le q_j(p^\mathsf{T}B)_j = \sum_k b_{kj}p_kq_j = \sum_k b_{kj}s_{kj}$$

From the remark in the solution of Exercise 2.1.2 we know that for a Nash Equilibrium (p,q) and constants $c,c'\in\mathbb{R}$ we have $(Aq)_i=c$ for all i such that $p_i\neq 0$, and symmetrically $(p^\mathsf{T}B)_j=c'$ for all j for which $q_j\neq 0$. Notice that $p\cdot(Aq)=c$ and that $q\cdot(p^\mathsf{T}B)=c'$ since p,q are probability vectors, and by the aforementioned property.

Now if $p_i=0$ then we trivially have $p_i(Aq)_{i'}=0=p_i(Aq)_i$, and similarly for $q_j=0$. Hence assume that $p_i,q_j\neq 0$. Then

$$p_i(Aq)_{i'} = p_i(e_{i'}\cdot Aq) \le p_i(p\cdot Aq) = p_ic = p_i(Aq)_i$$

by the definition of (p,q) being a Nash Equilibrium. Similarly

$$q_j(p^\mathsf{T}B)_{j'} = q_j(B^\mathsf{T}p\cdot e_{j'}) = q_j(p\cdot Be_{j'}) \le q_j(p\cdot Bq) = q_jc' = q_j(p^\mathsf{T}B)_j$$

as we claimed, where the inequality follows again by (p,q) is a Nash Equilibrium.

2) Consider the battle of the sexes game (coordination game) with bimatrix

$$\begin{pmatrix}(2,1) & (0,0)\\ (0,0) & (1,2)\end{pmatrix}.$$

Recall that in Exercise 2.4 we analysed all the possible phase diagrams for 2×2 replicator games. Using the same notation as in Exercise 2.4 we know that

$$\alpha_1=-1 \qquad \alpha_2=-2 \qquad \beta_1=-2 \qquad \beta_2=-1$$

which does confirm that we are considering a coordination game since $\alpha_1\alpha_2>0$, $\beta_1\beta_2>0$, and $\alpha_1\beta_1>0$. The Mixed Nash Equilibrium is represented on $I^2=[0,1]\times[0,1]$ by $\theta=\left(\frac{\beta_1}{\beta_1+\beta_2},\frac{\alpha_1}{\alpha_1+\alpha_2}\right)=\left(\frac{2}{3},\frac{1}{3}\right)$, hence in the phase space $\Delta_A\times\Delta_B$ it corresponds to $(p^A,p^B)=\left(\left(\frac{2}{3},\frac{1}{3}\right),\left(\frac{1}{3},\frac{2}{3}\right)\right)$. The same question immediately tells us that the other two (Pure) Nash Equilibria are given by $((0,1),(0,1))=(F,F)$, which corresponds to payoff $(1,2)$, and by $((1,0),(1,0))=(T,T)$ which corresponds to the payoff $(2,1)$.

Let us break down the precedent bimatrix into two matrices

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

and let P be a joint distribution. Therefore, we have

$$\begin{aligned}
2p_{21} = a_{11}p_{21} + a_{12}p_{22} = \sum_k a_{1k}p_{2k} &\leq \sum_k a_{2k}p_{2k} = a_{21}p_{21} + a_{22}p_{22} = p_{22} \\
p_{12} = a_{21}p_{11} + a_{22}p_{12} = \sum_k a_{2k}p_{1k} &\leq \sum_k a_{1k}p_{1k} = a_{11}p_{11} + a_{12}p_{12} = 2p_{11} \\
2p_{21} = b_{12}p_{11} + b_{22}p_{21} = \sum_k b_{k2}p_{k1} &\leq \sum_k b_{k1}p_{k1} = b_{11}p_{11} + b_{21}p_{21} = p_{11} \\
p_{12} = b_{11}p_{12} + b_{21}p_{22} = \sum_k b_{k1}p_{k2} &\leq \sum_k b_{k2}p_{k2} = b_{12}p_{12} + b_{22}p_{22} = 2p_{22}.
\end{aligned}$$

Since $2p_{21} \leq p_{22}$ and $2p_{21} \leq p_{11}$ if follows that $p_{21} \leq \frac{1}{2}\min(p_{11}, p_{22})$, and similarly since $p_{12} \leq 2p_{11}$ and $p_{12} \leq 2p_{22}$, it follows that $p_{12} \leq 2\min(p_{11}, p_{22})$.

Suppose that the joint distribution P is induced by one of the Nash Equilibria, then we want to numerically show that P is a Correlated Equilibrium. For example if we consider (p^A, p^B) then we have $p_{11} = p_1^A p_1^B = \frac{2}{9}$, $p_{12} = p_1^A p_2^B = \frac{4}{9}$, $p_{21} = p_2^A p_1^B = \frac{1}{9}$, $p_{22} = p_2^A p_2^B = \frac{2}{9}$. Clearly all the inequalities are respected

$$\begin{aligned}
\frac{1}{9} &= p_{21} \leq \frac{1}{2}\min(p_{11}, p_{22}) = \frac{1}{2}\frac{2}{9} = \frac{1}{9} \\
\frac{4}{9} &= p_{12} \leq 2\min(p_{11}, p_{22}) = 2\frac{2}{9} = \frac{4}{9}
\end{aligned}$$

hence the distribution induced by (p^A, p^B) is a CE.

If we now consider the Nash Equilibrium $(T, T) = ((1, 0), (1, 0))$, we have

$$p_{11} = 1, \qquad p_{12} = 0, \qquad p_{21} = 0, \qquad p_{22} = 0$$

which means that since $\min(p_{11}, p_{22}) = p_{22} = 0$, and $0 = p_{12} = p_{21} \leq \frac{1}{2}\min(p_{11}, p_{22}) = 0$ the probability distribution P induced by this Nash Equilibrium is a CE.

Finally, the last Nash Equilibrium we are left to check is $(F, F) = ((0, 1), (0, 1))$ which gives us the following entries for P

$$p_{11} = 0, \qquad p_{12} = 0, \qquad p_{21} = 0, \qquad p_{22} = 1.$$

Following the same argument as for (T, T) we have that the probability distribution induced by (F, F) is a CE, as we expected.

Therefore all the three probability distributions corresponding to the Nash Equilibria for this game are in the CE set. If we denote by $P_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ the distribution induced by the NE (T, T) and by $P_2 \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ the one induced by the NE (F, F) then we can see that for any $\sigma \in [0, 1]$ the matrix

$$\tilde{P} = \sigma P_1 + (1 - \sigma)P_2 = \begin{pmatrix} \sigma & 0 \\ 0 & 1 - \sigma \end{pmatrix}$$

is in the CE set, by simply checking the inequalities we have derived before. We can conclude that the CE set is infinite. Clearly if $\sigma = \frac{1}{2}$ then $\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$ an element of the CE set.

Consider the joint probability distribution described by $\tilde{P} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$. The expected payoff for the first player is given by summing up all the possible payoffs multiplied by the probability that such payoffs are achieved. Therefore

$$\mathbb{E}(\text{Payoff}_A|\tilde{P}) = 2\frac{1}{2} + 0 + 0 + 1\frac{1}{2} = \frac{3}{2}$$

since $((1,0),(1,0))$ is played with probability $\frac{1}{2}$ and it has payoff 2, whereas $((0,1),(0,1))$ is played with probability $\frac{1}{2}$ with payoff 1. A similar computation gives us that $\mathbb{E}(\text{Payoff}_B|\tilde{P}) = 1\frac{1}{2} + 2\frac{1}{2} = \frac{3}{2}$. The expected payoff for the distribution $\tilde{P}$ is $(\frac{3}{2}, \frac{3}{2})$.

The expected payoff for the Mixed Nash Equilibrium θ which induces a probability distribution $P_\theta = \begin{pmatrix} \frac{2}{9} & \frac{4}{9} \\ \frac{1}{9} & \frac{2}{9} \end{pmatrix}$ is given by

$$\begin{aligned} \mathbb{E}(\text{Payoff}_A|P_\theta) &= 2\frac{2}{9} + 0\frac{4}{9} + 0\frac{1}{9} + 1\frac{2}{9} = \frac{2}{3} \\ \mathbb{E}(\text{Payoff}_B|P_\theta) &= 1\frac{2}{9} + 0\frac{4}{9} + 0\frac{1}{9} + 2\frac{2}{9} = \frac{2}{3} \end{aligned}$$

so this Nash Equilibrium is outperformed by $\tilde{P}$.

Since $a_{12} = a_{21} = b_{12} = b_{21} = 0$, in order to maximise the expected payoff we need to look at probability distributions with shape $Q = \begin{pmatrix} \sigma & 0 \\ 0 & 1-\sigma \end{pmatrix}$ where $\sigma \in [0,1]$. If $\sigma = 1$ then $\mathbb{E}(\text{Payoff}_A|Q) = \mathbb{E}(\text{Payoff}_A|P_1)$ is maximised (and the total expected payoff is $(2,1)$), whereas for $\sigma = 0$ then $\mathbb{E}(\text{Payoff}_B|Q) = \mathbb{E}(\text{Payoff}_B|P_2)$ is maximised (and the expected payoff is $(1,2)$). In general, the expected payoff is given by $(2\sigma + (1-\sigma), \sigma + 2(1-\sigma)) = (\sigma + 1, 2-\sigma)$, which means that every σ gives a Pareto optimal expected payoff: any improvement to one of the player's payoff will negatively affect the other player payoff.

Next we want to show that playing the mixed Nash Equilibrium truly leads to the worst payoff. Consider a probability matrix of the form

$$\hat{P} = \begin{pmatrix} \frac{2}{9} + \varepsilon_1 & \frac{4}{9} + \varepsilon_2 \\ \frac{1}{9} + \varepsilon_3 & \frac{2}{9} + \varepsilon_4 \end{pmatrix}$$

where $\sum_i \varepsilon_i = 0$ and

$$\begin{aligned} -\frac{4}{9} \leq \varepsilon_2 \leq 2\min(\varepsilon_1, \varepsilon_4) \\ -\frac{1}{9} \leq \varepsilon_3 \leq \frac{1}{2}\min(\varepsilon_1, \varepsilon_4) \end{aligned}$$

since we want $\hat{P}$ to be a CE. Notice that ε_1 and ε_4 cannot be both negative, or that would imply that ε_2 and ε_3 are negative as well, contradicting the assumption that $\sum_i \varepsilon_i = 0$. If we turn our attention to the expected payoffs given $\hat{P}$ we will see that these amount to

$$\mathbb{E}(\text{Payoff}_A|\hat{P}) = \frac{2}{3} + 2\varepsilon_1 + \varepsilon_4$$
$$\mathbb{E}(\text{Payoff}_A|\hat{P}) = \frac{2}{3} + \varepsilon_1 + 2\varepsilon_4.$$

If both $\varepsilon_1, \varepsilon_4$ are positive then the expected payoff given $\hat{P}$ is higher than the one expected given P_θ. If we instead assume that ε_1 is positive, and ε_4 is negative then $\varepsilon_1 = -\varepsilon_2 - \varepsilon_2 - \varepsilon_3 > 0$ and

$$\mathbb{E}(\text{Payoff}_A|\hat{P}) = \frac{2}{3} + 2\varepsilon_1 + \varepsilon_4 = \frac{2}{3} - 2\varepsilon_1 - 2\varepsilon_2 - \varepsilon_4 > \frac{2}{3}$$
$$\mathbb{E}(\text{Payoff}_A|\hat{P}) = \frac{2}{3} + \varepsilon_1 + 2\varepsilon_4 = \frac{2}{3} - \varepsilon_1 - \varepsilon_2 + \varepsilon_4 > \frac{2}{3} - 2\varepsilon_4 - \frac{1}{2}\varepsilon_4 + \varepsilon_4 > \frac{2}{3},$$

which confirms that this would give a better payoff than playing according to P_θ. Notice that assuming ε_1 negative, and ε_4 positive leads to the same result given the symmetry in the coefficients of ε_1 and ε_4 in the functions $\mathbb{E}(\text{Payoff}_A|\hat{P})$, and $\mathbb{E}(\text{Payoff}_A|\hat{P}$. We can conclude that playing according to P_θ is always leading to the worse payoff.

Given the Pareto optimality for $\tilde{P}$ and the payoffs for $\hat{P}$, we can conclude that whenever the intermediator recommends playing according to a distribution P in the CE set for which p_{12} and p_{21} are not simultaneously zero, then the payoffs are not maximised. Indeed, if the intermediator recommends strategies such as (T, F) or (F, T) then no players will see any benefit in terms of payoff to such strategy. On the other hand, if they recommend to play according to matrices like $\tilde{P}$ then is clear that the intermediator in this game will always favour one particular player over the other (unless $\sigma = \frac{1}{2}$ then both players will be treated in the same way and get the same payoff).

3) Consider the game of chicken described by a bimatrix

$$\begin{pmatrix} (6,6) & (2,7) \\ (7,2) & (0,0) \end{pmatrix}$$

which can be decomposed into

$$A = \begin{pmatrix} 6 & 2 \\ 7 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 6 & 7 \\ 2 & 0 \end{pmatrix}.$$

As we did for the previous exercise, we can determine the phase portrait of this two players 2×2 game by looking at the coefficients α, β as defined in Exercise 2.4. If we adopt the second convention, we have

$$\alpha_1 = 2, \qquad \alpha_2 = 1, \qquad \beta_1 = 2, \qquad \beta_2 = 1$$

which translates to $\alpha_1\alpha_2 > 0$, $\beta_1\beta_2 > 0$, and $\alpha_1\beta_1 > 0$. Therefore, the game of chicken is a coordination game. We have three Nash Equilibria: two pure, and one mixed. The mixed Nash Equilibria is given by $\theta = \left(\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{1}{3}\right)\right)$, whereas the two Pure Nash Equilibria $(D, C) = ((0,1), (1,0))$, and $(C, D) = ((1,0), (0,1))$.

As before we want to show some CE inequalities. These are obtain through the following computations

$$\begin{aligned}
6p_{21} + 2p_{22} &= \sum_k a_{1k}p_{2k} \leq \sum_k a_{2k}p_{2k} = 7p_{21} \\
7p_{11} &= \sum_k a_{2k}p_{1k} \leq \sum_k a_{1k}p_{1k} = 6p_{11} + 2p_{12} \\
6p_{12} + 2p_{22} &= \sum b_{k1}p_{k2} \leq \sum_k b_{k2}p_{k2} = 7p_{12} \\
7p_{11} &= \sum_k b_{k2}p_{k1} \leq \sum_k b_{k1}p_{k1} = 6p_{11} + 2p_{21}.
\end{aligned}$$

These inequalities can be rewritten as

$$p_{22} \leq \frac{1}{2}\min(p_{12}, p_{21}) \qquad p_{11} \leq 2\min(p_{12}, p_{21}).$$

If we consider the probability distribution $P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix}$ then we immediately see that

$$\begin{aligned}
0 = p_{22} &< \frac{1}{2}\min(p_{12}, p_{21}) = \frac{1}{6} \\
\frac{1}{3} = p_{11} &< 2\min(p_{12}, p_{21}) = \frac{2}{3}
\end{aligned}$$

so P is in the Correlated Equilibrium set. The expected payoff when playing $P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix}$ is given by

$$\begin{aligned}
\mathbb{E}(\text{Payoff}_A | P) &= \frac{1}{3}6 + \frac{1}{3}2 + \frac{1}{3}7 = 5 \\
\mathbb{E}(\text{Payoff}_B | P) &= \frac{1}{3}7 + \frac{1}{3}2 + \frac{1}{3}6 = 5.
\end{aligned}$$

On the other hand, the probability distribution induced by the Nash Equilibrium $(D, C) = ((0,1), (1,0))$ is given by $P_1 = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ with payoff $(7, 2)$, and similarly the probability distribution induced by the Nash Equilibrium $(C, D) = ((1,0), (0,1))$ is given by $P_2 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ with payoff $(2, 7)$. The Mixed Nash Equilibrium $\theta = \left(\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{1}{3}\right)\right)$ induces a probability distribution $P_\theta = \begin{pmatrix} \frac{4}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{1}{9} \end{pmatrix}$ with payoff $\left(\frac{14}{3}, \frac{14}{3}\right)$. Playing according to P allows for a higher payoff for both players than playing P_θ, and this Correlated Equilibrium does not advantage a player over the other.

Indeed, when playing P the trusted intermediator is recommending both players to play strategies that generate positive payoff with equal probabilities. The intermediator is not siding with any player (the expected payoff is the same for both parties) and they are only strongly discouraging the two players to play (D,D) given its $(0,0)$ payoff.

7.2 Hart and Mas-Colell's regret matching

Exercise 7.2:
Consider the battle of the sexes game with bimatrix

$$\begin{pmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{pmatrix}$$

where the first action corresponds to watching Football, and the second one to watching Tennis. Figure 15 in the notes represents the behaviour of the four functions $\mathrm{DIFF}^t_A(F,T), \mathrm{DIFF}^t_A(T,F), \mathrm{REGRET}^t_A(F,T)$, and $\mathrm{REGRET}^t_A(T,F)$ With f^t reported along the x-axis, and the values of the functions along the y-axis . All these functions intersect at $f^t = \frac{2}{3}$ (and they are all zero at that point). Recall that $f^t = f^t_B(T)$, so we have that $\begin{pmatrix} f^t_B(F) \\ f^t_B(T) \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$. This corresponds to the second component of the the interior Nash Equilibrium of this game, namely $\left(\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{2}{3}\right)\right)$. Whenever Player 2 chooses to play a mixed strategy where the proportion of times they play T or F is determined by the internal Nash Equilibrium of B, then player 1 has no preferred way of replying to the strategies of player 2, since $\mathrm{REGRET}^t_A(F,T) = 0$, and $\mathrm{REGRET}(T,F)^t_A(T,F) = 0$. This is in accordance with the idea that the Nash Equilibrium is somewhat optimal for player 2.

We now want to show that for $t \geq 1$ we have $|f^{t+1} - f^t| < \frac{1}{t}$. By definition we see that

$$\begin{aligned} f^{t+1} &= \frac{1}{t+1}\#\{1 \leq i \leq t+1 \,|\, y^i = T\} \\ &\leq \frac{1}{t+1}(\#\{1 \leq i \leq t \,|\, y^i = T\} + 1) = \frac{t}{t+1}f^t + \frac{1}{t+1} \end{aligned}$$

and similarly

$$\begin{aligned} f^{t+1} &= \frac{1}{t+1}\#\{1 \leq i \leq t+1 \,|\, y^i = T\} \\ &\geq \frac{1}{t+1}\#\{1 \leq i \leq t \,|\, y^i = T\} = \frac{t}{t+1}f^t \end{aligned}$$

therefore

$$f^{t+1} - f^t \leq \frac{t}{t+1}f^t + \frac{1}{t+1} - f^t = \frac{1}{t+1} - \frac{1}{t+1}f^t \leq \frac{1}{t+1}$$

and

$$f^{t+1} - f^t \geq \frac{t}{t+1}f^t - f^t = -\frac{f^t}{t+1} \geq -\frac{1}{t+1}$$

since $f^t \in [0,1]$. We can rewrite the last inequalities more compactly as

$$\left|f^{t+1} - f^t\right| \leq \frac{1}{t+1} < \frac{1}{t}.$$

Notice this condition is too weak: it does not allow us to say that the sequence of frequencies $(f^t)_t$ tend to any limit. Indeed, it can happen that the frequency arbitrarily oscillates between 0 and 1. This has an impact on no-regret algorithm. Such algorithm is based on looking at the probabilities p_j^{t+1} and $p_{j^*}^{t+1}$ in order to make a decide on what is the best move to play, and in our case we can see that these probabilities depend on the frequencies f^t (REGRET depends on f^t). Since $(f^t)_t$ does not need to converge, neither do the sequences $(p_j^t)_t$ and $(p_{j^*}^t)_t$. This means that we do not have to reach a point in our game where playing one specific strategy will be the answer to minimising regret. This situation might appear seem quite bleak, but there is an upside to this whole situation.

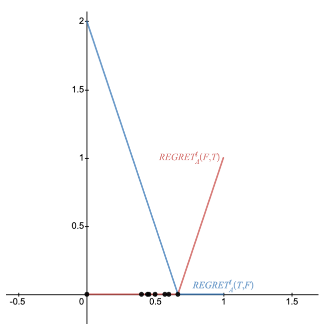


Figure 24: Player A regret when Player B chooses strategy T at prime times (11 time steps).

By Hart and Mas-Colell Theorem (Theorem 7.1 in the Lecture Notes) we know that if a player follows the no-regret algorithm, then they will (almost surely) asymptotically get zero regret for their moves, this means that even if the probabilities p_j^{t+1} and $p_{j^*}^{t+1}$ can vary depending on the how the second player decides to play, if one sticks to what they recommend, then it is (almost) always possible to get very little regret on the long run. One can think of this algorithm as dynamically adapting itself with respect to the second player choice of strategies.

Let us take a look at a more numerical examples, hoping that it will make the whole discussion clearer. Suppose that player B plays strategy T at prime times then the first few rounds of this game will look something like

Time t	1	2	3	4	5	6	7	8	9	10	11	...
Strategy$_B$	F	T	T	F	T	F	T	F	F	F	T	...
Frequency f^t	0	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{1}{2}$	$\frac{4}{7}$	$\frac{1}{2}$	$\frac{4}{9}$	$\frac{2}{5}$	$\frac{5}{11}$	...
Recom. strategy$_A$	F	F	T/F	F	F	F	F	F	F	F	F	...

Table 4: No-Regret algorithm for when player B plays T at prime times.

What is happening in this case is that player B always plays F, and occasionally they play strategy T, at random (since it is not possible to predict when the next prime number will appear, for time large enough). The no-regret algorithm does not get thrown off from this sporadic appearance of strategy T from player B, but instead it immediately updates the probabilities $p_{j^*}^{t+1}$ and p_j^{t+1} so that player A can minimise regret after what could be considered an unexpected action.

We have just analysed what happens when player B mostly sticks to one strategy, and plays the other one at random times. We can similarly look at the case where player B sticks for a long time to playing one strategy, and at some random time switches to the other strategy, and keeps playing it for an even longer amount of time. One way to model this is to consider an increasing sequence of real numbers $(n_i)_i$ tending to infinity such that $n_{i+1} \geq n_i^2$, and then letting player B play strategy F (resp. T) when i is even (resp. odd) for times $e^{n_i}, \dots, e^{n_{i+1}-1}$. Notice that

$$\lim_{n\to\infty} \frac{e^{n_{i+1}} - e^{n_i}}{e^{n_i}} = \lim_{n\to\infty} e^{n_{i+1}-n_i} - 1 \geq \lim_{n\to\infty} e^{n_i^2 - n_i} - 1 = +\infty$$

which roughly tells us that the amount of times B will play a different strategy from before is quite considerable (strategies are not sporadically play as before). The same argument as before holds even in this case: the no-regret algorithm adapts so that the regret is going to tend to zero almost surely.

As a final remark, please notice that the Hart and Mas-Colell algorithm does not take into account the payoff for the second player, it is only interested in the payoff of the player following the algorithm. As explained before, following this algorithm will almost surely minimise the regret connected your to choices, this algorithm does not try to minimise the payoff of your opponent, or to maximise their regret.

7.3 Min-max solutions and zero-sum games

Exercise 7.2:
1) Consider the zero-sum game described by the bimatrix

$$G = \begin{pmatrix} (4,-4) & (-2,2) \\ (-5,5) & (6,-6) \end{pmatrix}.$$

and in order to simplify notation, we will denote the first strategy by T and the second one by F. The expected payoffs for Alice (player 1) against Bob (player 2) whenever she plays a random strategy with probability $(p, 1-p)$ is

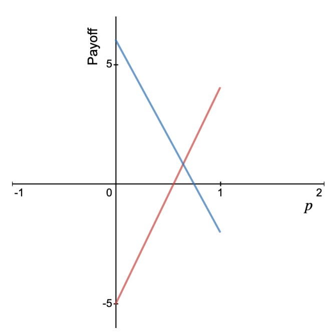


Figure 25: Expected Payoff for Alice playing against Bob 1^{st} strategy (in red) and and 2^{nd} strategy (in blue).

given by $\mathbb{E}(\text{Payoff}_A \mid \text{Bob plays strategy 1}) = \mathbb{E}(\text{Payoff}_A \mid (\cdot, T)) = 9p - 5$ and $\mathbb{E}(\text{Payoff}_A \mid \text{Bob plays strategy 2}) = \mathbb{E}(\text{Payoff}_A \mid (\cdot, F)) = 6 - 8p$. These two functions are represented in Figure 25, and we have that they meet at the point $p = \frac{11}{17}$, where Alice would expect a payoff of $\frac{14}{17}$ against Bob. This payoff is independent from Bob's choice of strategy. We can repeat the same analysis for Bob. Suppose that Bob plays strategy 1 with probability q, and strategy 2 with probability $1 - q$ then their expected payoffs are given by $\mathbb{E}(\text{Payoff}_B|(T, \cdot)) = -4q+2(1-q) = 2-6q$ if Alice plays T, and $\mathbb{E}(\text{Payoff}_B|(F, \cdot)) = 5q+-6(1-q) = 11q - 6$ if Alice plays F. The two expected payoffs are given by two lines meeting at $q = \frac{8}{17}$. As before, if $q = \frac{8}{17}$, then Bob can expect a payoff of $-\frac{14}{17}$, independently from Alice's strategy.

2) Consider the zero-sum game associated to the matrix

$$A = \begin{pmatrix} 4 & 1 & -4 \\ 3 & 2 & 5 \\ 0 & 1 & 7 \end{pmatrix}$$

where we assume the second convention. As explained in the text of the exercise, the entry $a_{22} = 2$ is a *saddle-point* of the game since it is the biggest entry in its column, and the smallest in its row. This property implies that the pure strategy (e_2, e_2) is a Nash Equilibrium.

Recall that in the Lecture Notes you studied that a point $(\hat{x}, \hat{y})$ is a Nash Equilibrium for a zero sum game defined by a matrix A if

$$\min_y \hat{x} \cdot Ay = v = \max_x x \cdot A\hat{y}.$$

Given a saddle point a_{ij} then the product $e_i \cdot Ay$ is equal to $(e_i^\intercal A)y$ and $e_i^\intercal A$ corresponds to the i^{th} row of A, and similarly if we look at the product $x \cdot Ae_j$ then Ae_j corresponds to the j^{th} column of A. Since we have assumed that a_{ij} is the smallest element in its row, then $\min_y e_i^\intercal Ay = a_{ij}$, and given that we assumed that it was also the biggest entry in its column then $\max_x x \cdot Ay = a_{ij}$, which by the result we recalled from the lecture notes exactly means that (e_i, e_j) is a Nash Equilibrium.

Let us look at a more concrete example. For the matrix A reported above we have

$$\min_y e_2 \cdot Ay = \min_y \begin{pmatrix} 3, & 2, & 5 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = 2$$

$$\max_x x \cdot Ae_2 = \max_x \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = 2$$

which confirms that (e_2, e_2) is a Nash Equilibrium for the zero-sum two player games given by the matrices $(A, -A)$.

7.4 Another way of thinking of the min-max theorem

Exercise 7.4:
1) We wish to prove the following proposition.

Proposition 7.1. *Given a function $A : \Delta \times \Delta \to \mathbb{R}$ then* $\min_p \max_q A(p, q) = \max_q \min_p A(p, q)$ *if and only if for any v in the image of A*

i) $\exists p, \forall q$ *such that* $A(p, q) \leq v$

ii) $\forall q, \exists p$ *such that* $A(p, q) \leq v$.

We will prove the two directions of the "if and only if" implication.

$\Longleftarrow$) In order to reach a contradiction, suppose that the min-max equality $\min_p \max_q A(p, q) = \max_q \min_p A(p, q)$ does not hold, whilst $i)$ and $ii)$ are equivalent (either both true or both false). Notice that

$$\min_{p(q)} A(p, q) \leq A(p, q) \leq \max_{q(p)} A(p, q)$$

where the LHS is independent of p (and the RHS is independent of q). By minimising p throughout the inequality we get

$$\min_{p(q)} A(p, q) \leq \min_p \max_{q(p)} A(p, q)$$

and if we now maximise q we have

$$\max_q \min_{p(q)} A(p, q) \leq \min_p \max_{q(p)} A(p, q).$$

Since we have assumed that the min-max equality does not hold, we have the strict inequality

$$\max_q \min_{p(q)} A(p, q) < \min_p \max_{q(p)} A(p, q).$$

and therefore there exists $v \in \mathbb{R}$ so that

$$\max_q \min_{p(q)} A(p, q) < v < \min_p \max_{q(p)} A(p, q). \tag{13}$$

Since $\max_q \min_{p(q)} A(p, q) < v$, this means that for all q, there exists p so that $A(p, q) < v$. Since p, q are taken from a compact domain, there exists $v' < v$ ($v' := \max_q \min_{p(q)} A(p, q)$) such that for all q there exists a p for which $A(p, q) \leq v'$. At the beginning of this proof we assumed that $i)$ and $ii)$ are equivalent, and we have just showed that $ii)$ holds, therefore $i)$ has to hold as well: there exists a p such that $A(p, q) \leq v'$ for all q. This last statement is equivalent to $\min_p \max_{q(p)} A(p, q) \leq v' < v$, which contradicts inequality 13.

$\Longrightarrow$) This implication is reached by noticing that minimisation corresponds to the existential quantifier and maximisation corresponds to the universal quantifier.

We now want to show that if we assume $\min_p \max_q A(p,q) = \max_q \min_p A(p,q)$ then both conditions $i)$ and $ii)$ are either both true or both false, i.e. they are equivalent. In order to reach a contradiction, suppose there exists $v \in \mathbb{R}$ so that $i)$ holds, and $ii)$ does not. The first condition being true yields $\min_p \max_{q(p)} A(p,q) \leq v$. On the other hand, we have (by negating $ii)$): $\exists q^*$ such that $\forall p$ we have $A(p,q^*) > v$. Hence $\min_p A(p,q^*) > v$, which implies

$$\max_q \min_{p(q)} A(p,q) \geq \min_p A(p,q^*) > v.$$

By applying the min-max equality we end up with

$$v \geq \min_p \max_{q(p)} A(p,q) = \max_q \min_{p(q)} A(p,q) > v$$

which is a contradiction.

Let us now assume there exists $v \in \mathbb{R}$ so that $ii)$ holds and that $i)$ does not. Now $ii)$ yields $\max_q \min_{p(q)} A(p,q) \leq v$. The negation of $i)$ translates to $\forall p$ then $\exists q^*$ such that $A(p,q^*) > v$, which means that $\max_{q(p)} A(p,q) > v$ for any p, and in particular $\min_p \max_{q(p)} A(p,q) > v$. As before, this leads to a contradiction.

We can conclude that if the min-max equality holds then either $i)$ and $ii)$ are either both true or both false.

2) Consider the function

$$\begin{aligned} A : [0,1] \times [0,1] &\to [0,1] \\ (p,q) &\mapsto pq \end{aligned}$$

and we want to show $\max_q \min_p A(p,q) = \min_p \max_q A(p,q)$. For such a simple function this is quite straightforwards: the minimum of pq for either p or q is always 0, independently from the value of the other variable (this is only true since we work over $[0,1] \times [0,1]$. Therefore $\max_q \min_p A(p,q) = \min_p \max_q A(p,q) = 0$.

We wish to show that the equivalent statement of the min-max theorem stated at the beginning of this question holds for this toy model. Since A maps onto $[0,1]$ then choosing $v < 0$ makes no sense. Fix $v \geq 0$, then fix $p \leq v$ then we have that for all $q \in [0,1]$

$$A(p,q) = pq \leq p \leq v \qquad\qquad \text{since } q \leq 1, \text{ and } p \leq v$$

therefore $i)$ holds. Similarly, for $v \geq 0$ fixed and any $q \in [0,1]$ then choose $p \geq v$ then

$$A(p,q) = pq \leq p \leq v \qquad\qquad \text{since } q \leq 1, \text{ and } p \leq v$$

hence $ii)$ holds as well, as we expected.

3) Consider now the function

$$
\begin{aligned}
A : [0,1] \times [0,1] &\to \mathbb{R} \\
(p,q) &\mapsto \begin{cases} p+q & \text{if } p+q \leq 1 \\ 2-(p+q) & \text{otherwise.} \end{cases}
\end{aligned}
$$

Firstly we can see that

$$
A(p,q) \leq \begin{cases} 1 & \text{if } p+q \leq 1 \\ 2-1 & \text{if } p+q \geq 1 \end{cases} = 1.
$$

hence $\max_q A(p,q) = 1$ (equality is reached, for example, when $q = 1-p$ for any $p \in [0,1]$).

Similarly, if we try to minimise $A(p,q)$ for p we get

$$
A(p,q) \geq \begin{cases} 0+q & \text{if } 0+q \leq 1 \\ 2-(1+q) & \text{if } 1+q \geq 1 \end{cases} = \begin{cases} q & \text{if } q \leq 1 \\ 1-q & \text{if } q > 0 \end{cases}
$$

where the first function is minimised for $p = 0$, and the second one for $p = 1$. Therefore, we can write more compactly $\min_p A(p,q) = \min(q, 1-q)$. Now the minmax theorem does not hold anymore since

$$
\begin{aligned}
\min_p \max_q A(p,q) &= \min_p(1) = 1 \\
\max_q \min_p A(p,q) &= \max_q \min(q, 1-q) = \frac{1}{2}.
\end{aligned}
$$

Some convexity and concavity properties on the function $A(p,q)$ are required in order for the minmax theorem to hold. Let M, and N be two subsets of a topological vector spaces $\mathfrak{U}$, and $\mathfrak{V}$ (these are just vector spaces equipped with a topology so that vector addition and scalar multiplication are continuous with respect to the chosen topology). Assume the scalar field to be either $\mathbb{R}$ or $\mathbb{C}$ equipped with the Euclidean (or Standard) topology.

Definition 7.1. A function f on $M \times N$ is *quasi-concave* in N if $\{y \mid f(x,y) \geq c\}$ is a convex set for any $x \in M$ and $c \in \mathbb{R}$. Similarly, a function f on $M \times N$ is *quasi-convex* in M if $\{x \mid f(x,y) \leq c\}$ is a convex set for any $y \in N$ and $c \in \mathbb{R}$.

Theorem 7.1 (Sion's Minmax Theorem [2])**.** *Let M be a compact convex subset of $\mathfrak{U}$, and let N be a subset of $\mathfrak{V}$. If f is a real-valued function on $M \times N$ with*

- *$f(x,\cdot)$ upper semicontinuous and quasi-concave on N, $\forall x \in M$;*
- *$f(\cdot,y)$ lower semicontinuous and quasi-convex on M, $\forall y \in N$;*

then

$$
\min_M \sup_N f(x,y) = \sup_N \min_M f(x,y).
$$

[2]Sion, M. (1958) On general Minmax Theorems. *Pacific Journal of Mathematics.* 8(1), 171-176.

7.5 A vectored valued payoff game

Exercise 7.5: 1) Let Δ represent our canonical simplex and consider a function

$$A : \Delta \times \Delta \to \mathbb{R}^k.$$

Let $\mathcal{C}$ be a convex subset of $\mathbb{R}^k$, such that for each $q \in \Delta$ there exists a $p \in \Delta$ such that $A(p,q) \in \mathcal{C}$. In the question we are asked if we can infer that there exists a $p \in \Delta$ such that for all $q \in \Delta$ we have $A(p,q) \in \mathcal{C}$. As the hint suggests, this is not true if the range has dimension greater or equal to 2. Let us provide a counterexample.

Let $\Delta = [0,1]$, consider the function

$$\begin{aligned} A : \Delta \times \Delta &\to [0,1]^2 \subset \mathbb{R}^2 \\ (p,q) &\mapsto (p,q) \end{aligned}$$

and let $\mathcal{C}$ be the diagonal of $[0,1]^2$, i.e. $\mathcal{C} \coloneqq \{(x,x) \mid x \in [0,1]\}$. Since $\mathcal{C}$ is an interval it is automatically convex. In this case, for any $q \in [0,1]$ set $p \coloneqq q \in [0,1]$ then $A(p,q) = (p,q) = (q,q) \in \mathcal{C}$. Unfortunately, there exists no p for which $A(p,q) \in \mathcal{C}$ for all q. Suppose such a p existed, then take $q = p + \frac{1}{4} \mod 1$. In this situation $A(p,q) = (p,q) = (p, p + \frac{1}{4} \mod 1)$ which is clearly not contained in $\mathcal{C}$. You can think of our choice of q as a vertical translation of $\mathcal{C}$ by $\frac{1}{4}$ in the two dimensional torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.

2) Now, let $\Delta = [0,1]$ and consider the function

$$\begin{aligned} A : \Delta \times \Delta &\to \mathbb{R} \\ (p,q) &\mapsto pq. \end{aligned}$$

We will say that a convex set $\mathcal{C}$ is *acceptable* if for all q there exists a p such that $A(p,q) \in \mathcal{C}$. For this particular choice of A, we have that $\mathcal{C}$ is acceptable if and only if it contains the point 0.

Since the set $\mathcal{C}$ is a convex subset of $\mathbb{R}$, this restricts its shape. Indeed $\mathcal{C}$ can only be a singleton, the whole real line $\mathbb{R}$, or (potentially unbounded) interval. Notice that $\text{Im } A = [0,1]$, hence if $\mathcal{C} \cap \text{Im } A = \emptyset$ then $\mathcal{C}$ is automatically not acceptable. In order to reach a contradiction let us assume that $\text{Im } A \cap \mathcal{C} \neq \emptyset$, but $0 \notin \mathcal{C}$. Let $a = \min \mathcal{C} > 0$, then if we take $q = 0$ for example, there are no $p \in [0,1]$ for which $A(p,q) = pq = 0$ could possible be greater than the positive number a. This means that for $\mathcal{C}$ to be acceptable it must contain 0.

Let us show that this is actually sufficient. If $0 \in \mathcal{C}$ then for any q let $p = 0$, and then $A(p,q) = pq = 0 \in \mathcal{C}$, as we claimed. A similar proof now gives us that if $\mathcal{C}$ is an acceptable set, then there exists $p \in \Delta$ such that for all $q \in \Delta$, we have that $A(p,q) \in \mathcal{C}$. If we fix $p = 0$, then for all $q \in \Delta$ we have that $A(p,q) = pq = 0 \in \mathcal{C}$, since $\mathcal{C}$ is acceptable.

7.6 Blackwell approachability theorem

Exercise 7.6:

1) In Figure 26 we have a diagram illustrating all the various vectors, sets, and

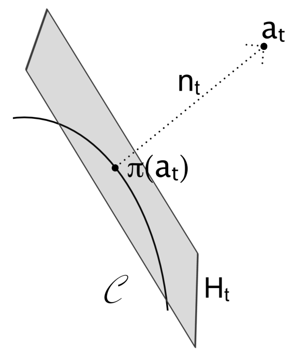


Figure 26: Schematic of the proof of the Blackwell Approachability Theorem

hyperplanes involved in the proof of Blackwell approachability theorem. Note that we only drew a part of the boundary of the convex set $\mathcal{C}$ in Figure 26.

We will now write down an algorithm on how to find p^t.

1. Consider the projection $\pi(a_t)$ on the convex set $\mathcal{C}$, where a_t is a vector representing the time average of the vector-valued payoff $A(p,q)$ up to time t;

2. Compute the vector $n_t = a_t - \pi(a_t)$, normal to $\mathcal{C}$, starting at πa_t and pointing towards a_t;

3. Define the half space $H_t \coloneqq \{a \mid a \cdot n_t \leq \pi(a_t) \cdot n_t\}$ which contains $\mathcal{C}$. Notice that $\{a \cdot n_t = \pi(a_t) \cdot n_t\}$ is a (hyper)plane passing though $\pi(a_t)$, perpendicular to n_t;

4. Rewrite $A(p,q) \cdot n_t$ as $p \cdot A^t q$, where A^t is a matrix depending on t;

5. By the approachability of H_t we have that the min-max theorem holds for $A(p,q) \cdot n_t$, hence there exists a p^t such that $A(p^t,q) \cdot n_t \leq \pi(a_t) \cdot n_t$ for all q.

This algorithm does not give you a way to explicitly compute v^t, but there exist plenty of packages in Python/Julia/Matlab which have been developed for to optimise the construction of such vector.

2) Assume that

$$a_t = \frac{1}{t} \sum_{i=1}^{t} A(p^i, q^i)$$

belongs to the convex set $\mathcal{C}$. As we have seen in the proof of the Blackwell Approachability Theorem

$$a_{t+1} = \frac{t}{t+1} a_t + \frac{1}{t+1} A(p^{t+1}, q^{t+1}).$$

Note that a_{t+1} is the convex combination of the vectors a_t and $A(p^{t+1}, q^{t+1})$. Clearly, if $A(p^{t+1}, q^{t+1}) \in \mathcal{C}$, we have by convexity of $\mathcal{C}$ that a_{t+1} is in $\mathcal{C}$. Unfortunately this is not always the case. Consider the point $m := \partial C \cap \langle a_t, A(p^{t+1}, q^{t+1})\rangle$, and assume that

$$|A(p^{t+1}, q^{t+1}) - m| > \frac{t}{t+1}|A(p^{t+1}, q^{t+1}) - a_t|.$$

Under this assumption we can conclude that $a_{t+1} \in \langle m, A(p^{t+1}, q^{t+1})\rangle$, and therefore outside of $\mathcal{C}$. This counterexample is quite abstract. Try to come up with your own counterexample (maybe let $\mathcal{C}$ be a point...).

7.7 Regret minimisation

Exercise 7.7:
1) Let us look at a specific game to discuss this question. Consider the "battle of the sexes" game given by

$$M = \begin{pmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{pmatrix}$$

and suppose that player 2 decides their play after having seen the strategy chosen by player 1. Let T, and F denote the two strategies for this game, respectively e_1 and e_2. Furthermore, suppose that player 2 holds a grudge against player 1 and so they always play the opposite strategy to 1, i.e. if player 1 plays T, then player 2 replies with F, and vice-versa. This leads to payoff 0 for both players at all times. In particular this leads to at least one of the regrets for player 1 being always strictly positive, which is clearly contradicting the first Hart and Mas-Colell Theorem (Theorem 7.1 in the lecture notes).

Similarly we can show how Theorem 7.2 in the notes will not holds in this situation. Let us disregard the fact that player 2 is definitely not following the no-regret algorithm, and let us focus on what the matrix of frequencies the conjoint actions of player 1 and 2 looks like. Given the choice of strategies from player 2 our probability distributions keeping track of the frequency of actions up to time t will only have weights on the off-diagonal terms

$$P_t = \begin{pmatrix} 0 & 1 - f^t \\ f^t & 0 \end{pmatrix}$$

where $f^t = f^t_B(T)$ the frequency at which player 2 plays T, the first of the available strategies. From Exercise 7.1 we have that P_t does not belong to the CE set for this game, for any t, as we claimed.