Chapter 5

Potential Games

5.1 Best response dynamics

The existence of a Nash equilibrium is clearly a desirable property of a strategic game. In this chapter we discuss some natural classes of games that do have a Nash equilibrium. First, notice the following obvious nondeterministic algorithm, called ***best response dynamics***, to find a Nash equilibrium (NE):

choose $s \in S_1 \times \cdots \times S_n$;
while s is not a NE **do**
  choose $i \in \{1, \ldots, n\}$ such that s_i is not a best response to s_{-i};
  $s_i :=$ a best response to s_{-i}
od

Obviously, this procedure does not need to terminate, for instance when the Nash equilibrium does not exist. Even worse, it may cycle when a Nash equilibrium actually exists. Take for instance the following extension of the Matching Pennies game already considered in Section 4.1:

	H	T	E
H	$1, -1$	$-1, \ 1$	$-1, -1$
T	$-1, \ 1$	$1, -1$	$-1, -1$
E	$-1, -1$	$-1, -1$	$-1, -1$

Then an execution of the best response dynamics may end up in a cycle

$$((H, H),\ (H, T),\ (T, T),\ (T, H))^*.$$

However, for various games, for instance the Prisoner's Dilemma game (also for n players) and the Battle of the Sexes game all executions of the best response dynamics terminate. This is a consequence of a general approach that forms the topic of this chapter.

First, note the following simple observation to which we shall return later in the chapter.

Note 14 (Best Response Dynamics) *Consider a strategic game for n players. Suppose that every player has a strictly dominant strategy. Then all executions of the best response dynamics terminate after at most n steps and their outcome is unique.*

Proof. Each strictly dominant strategy is a unique best response to each joint strategy of the opponents, so in each execution of the best response dynamics every player can modify his strategy at most once. □

5.2 Potentials

In this section we introduce the main concept of this chapter. Given a game $G := (S_1, \ldots, S_n, p_1, \ldots, p_n)$ we call the function $P : S_1 \times \cdots \times S_n \to \mathbb{R}$ a ***potential*** for G if

$$\begin{array}{l} \forall i \in \{1, \ldots, n\}\ \forall s_{-i} \in S_{-i}\ \forall s_i, s_i' \in S_i \\ p_i(s_i, s_{-i}) - p_i(s_i', s_{-i}) = P(s_i, s_{-i}) - P(s_i', s_{-i}). \end{array}$$

We call then a game that has a potential a ***potential game***.

The intuition behind the potential is that it tracks the changes in the payoff when some player deviates, without taking into account which one. The following observation explains the interest in the potential games.

Note 15 (Potential) *For finite potential games all executions of the best response dynamics terminate.*

Proof. At each step of each execution of the best response dynamics the potential strictly increases. □

Consequently, each finite potential game has a Nash equilibrium. This is also a consequence of the fact that by definition each maximum of a potential is a Nash equilibrium.

A number of games that we introduced in the earlier chapters are potential games. Take for instance the Prisoner's Dilemma game for n players from Example 3. Indeed, we noted already that in this game we have $p_i(0, s_{-i}) - p_i(1, s_{-i}) = 1$. This shows that $P(s) := n - \sum_{j=1}^{n} s_j$ is a potential function. Intuitively, this potential counts the number of players who selected 0, i.e., the defect strategy.

Also, the Battle of the Sexes is a potential game. We present the game and its potential in Figure 5.1.

	F	B
F	$2,1$	$0,0$
B	$0,0$	$1,2$

	F	B
F	2	1
B	0	2

Figure 5.1: The Battle of the Sexes game and its potential

Example 12 It is less trivial to show that the Cournot competition game from Example 6 is a potential game. Recall that the set of strategies for each player is $\mathbb{R}_+$ and payoff for each player i is defined by

$$p_i(s) := s_i(a - b\sum_{j=1}^{n} s_j) - cs_i$$

for some given a, b, c, where (these conditions play no role here) $a > c$ and $b > 0$.

We prove that

$$P(s) := a\sum_{i=1}^{n} s_i - b\sum_{i=1}^{n} s_i^2 - b\sum_{1\leq i<j\leq n} s_i s_j - \sum_{i=1}^{n} cs_i$$

is a potential.

To show it we use the fundamental theorem of calculus that states the following. If $f : [a, b] \to \mathbb{R}$ is a continuous function defined on a real interval $[a, b]$ and F is an antiderivative of f, then

$$F(b) - F(a) = \int_a^b f(t)\, dt.$$

Applying this theorem to the functions p_i and P we get that for all $i \in \{1, \ldots, n\}$, $s_{-i} \in S_{-i}$ and $s_i, s_i' \in S_i$ we have

$$p_i(s_i, s_{-i}) - p_i(s_i', s_{-i}) = \int_{s_i'}^{s_i} \frac{\partial p_i}{\partial s_i}(t, s_{-i})\, dt$$

and

$$P(s_i, s_{-i}) - P(s_i', s_{-i}) = \int_{s_i'}^{s_i} \frac{\partial P}{\partial s_i}(t, s_{-i})\, dt.$$

So to prove that P is a potential it suffices to show that for all $i \in \{1, \ldots, n\}$

$$\frac{\partial p_i}{\partial s_i} = \frac{\partial P}{\partial s_i}.$$

But for all $i \in \{1, \ldots, n\}$ and $s \in S_1 \times \cdots \times S_n$

$$\frac{\partial p_i}{\partial s_i}(s) = (a - b\sum_{j=1}^{n} s_j) - bs_i - c = a - 2bs_i - b \sum_{j \in \{1, \ldots, n\} \setminus \{j\}} s_j - c = \frac{\partial P}{\partial s_i}(s).$$

Note that the fact that Cournot competion is a potential game does not automatically imply that it has a Nash equilibrium. Indeed, the set of strategies is infinite, so the Potential Note 15 does not apply. □

Exercise 10 Find a potential game that has no Nash equilibrium.
Hint. Analyze the game from Example 9. □

Exercise 11 Suppose that P_1 and P_2 are potentials for some game G. Prove that there exists a constant c such that for every joint strategy s we have $P_1(s) - P_2(s) = c$. □

The potential tracks the precise changes in the payoff function. We can relax this requirement and only track the sign of the changes of the payoff function. This leads us to the following notion.

Given a game $G := (S_1, \ldots, S_n, p_1, \ldots, p_n)$ we call the function $P : S_1 \times \cdots \times S_n \to \mathbb{R}$ an ***ordinal potential*** for G if

$$\begin{aligned} &\forall i \in \{1, \ldots, n\}\ \forall s_{-i} \in S_{-i}\ \forall s_i, s_i' \in S_i \\ &p_i(s_i, s_{-i}) - p_i(s_i', s_{-i}) > 0 \text{ iff } P(s_i, s_{-i}) - P(s_i', s_{-i}) > 0. \end{aligned}$$

As an example consider a modification of the Prisoner's Dilemma game and its ordinal potential given in Figure 5.2.

	C	D
C	$2,2$	$0,3$
D	$3,0$	$1,2$

	C	D
C	0	1
D	1	2

Figure 5.2: A game and its ordinal potential

Note that this game has no potential. Indeed every potential has to satisfy the following conditions:

$$\begin{aligned} P(C,C) - P(D,C) &= -1, \\ P(D,C) - P(D,D) &= -2, \\ P(D,D) - P(C,D) &= 1, \\ P(C,D) - P(C,C) &= 1, \end{aligned}$$

which implies $0 = -1$. So the notion of an ordinal potential is more general than that of a potential.

Exercise 12 Prove that

$$P(s) := s_1 s_2 \ldots s_n (a - b \sum_{j=1}^{n} s_j - c)$$

is an ordinal potential for the Cournot competition game introduced in Example 6 and analyzed in Example 12. □

An even more general notion is the following one. Given a game $G := (S_1, \ldots, S_n, p_1, \ldots, p_n)$ we call the function $P : S_1 \times \cdots \times S_n \to \mathbb{R}$ a ***generalized ordinal potential*** for G if

$$\begin{aligned} &\forall i \in \{1, \ldots, n\}\ \forall s_{-i} \in S_{-i}\ \forall s_i, s_i' \in S_i \\ &p_i(s_i, s_{-i}) - p_i(s_i', s_{-i}) > 0 \text{ implies } P(s_i, s_{-i}) - P(s_i', s_{-i}) > 0. \end{aligned}$$

As an example consider the game and its generalized ordinal potential given in Figure 5.3.

It is easy to check that this game has no ordinal potential. Indeed, every ordinal potential has to satisfy

$$P(T,L) < P(B,L) < P(B,R) < P(T,R).$$

But $p_2(T,L) = p_2(T,R)$, so $P(T,L) = P(T,R)$.

	L	R
T	$1,0$	$2,0$
B	$2,0$	$0,1$

	L	R
T	0	3
B	1	2

Figure 5.3: A game and its generalized ordinal potential

We now characterize the finite games that have a generalized ordinal potential. These are precisely the games for which the best response dynamics, generalized to better responses, always terminates. We first introduce the used concepts.

Fix a strategic game $G := (S_1, \ldots, S_n, p_1, \ldots, p_n)$. By a ***profitable deviation*** we mean a pair (s, s') of joint strategies such that $s' = (s'_i, s_{-i})$ for some s'_i and $p_i(s') > p_i(s)$. We say then that s'_i is a ***better response*** of player i to the joint strategy s. An ***improvement path*** is a maximal sequence of profitable deviations. We say that G has the ***finite improvement property*** (***FIP***), if every improvement path is finite. Finally, by an ***improvement sequence*** we mean a prefix of an improvement path. Obviously, if G has the FIP, then it has a Nash equilibrium.

We can now state the announced result.

Theorem 16 (FIP) *Every finite game has a generalized ordinal potential iff it has the FIP.*

In the proof below we use the following classic result.

Lemma 17 (König's Lemma) *Any finitely branching tree is either finite or it has an infinite path.* □

Proof. Consider an infinite, but finitely branching tree T. We construct an infinite path in T, that is, an infinite sequence

$$\xi : n_0 \; n_1 \; n_2 \; \ldots$$

of nodes such that, for each $i \geq 0$, n_{i+1} is a child of n_i. We define ξ inductively such that every n_i is the root of an infinite subtree of T. As n_0 we take the root of T. Suppose now that $n_0, \ldots, n_i$ are already constructed. By induction hypothesis, n_i is the root of an infinite subtree of T. Since T is finitely branching, there are only finitely many children $m_1, \ldots, m_n$ of n_i. At least

one of these children is a root of an infinite subtree of T, so we take n_{i+1} to be such a child of n_i. This completes the inductive definition of ξ. $\square$

Proof of the FIP Theorem 16.
($\Rightarrow$) Let P be a generalized ordinal potential. Suppose by contradiction that an infinite improvement path exists. Then the corresponding values of P form a strictly increasing infinite sequence. This is a contradiction, since there are only finitely many joint strategies.

($\Leftarrow$) Consider a branching tree the root of which has all joint strategies as successors and whose branches are the improvement paths. Because the game is finite this tree is finitely branching. By the assumption the game has the FIP, so this tree has no infinite paths. Consequently by König's Lemma this tree is finite and hence the number of improvement sequences is finite. Given a joint strategy s define $P(s)$ to be the number of improvement sequences that terminate in s. Then in the considered game $(S_1, \ldots, S_n, p_1, \ldots, p_n)$

$$\begin{aligned}&\forall i \in \{1, \ldots, n\}\ \forall s_{-i} \in S_{-i}\ \forall s_i, s'_i \in S_i\\ &p_i(s_i, s_{-i}) - p_i(s'_i, s_{-i}) > 0 \text{ implies } P(s_i, s_{-i}) - P(s'_i, s_{-i}) = 1,\end{aligned}$$

so P is a generalized ordinal potential. $\square$

5.3 Congestion games

We now study an important class of games that have a potential. Until now we associated with each player a payoff function p_i. An alternative is to associate with each player a ***cost function*** c_i. Then the objective of each player is to minimize the cost. Consequently, when the cost functions are used, a joint strategy s is a Nash equilibrium if

$$\forall i \in \{1, \ldots, n\}\ \forall s'_i \in S_i\ c_i(s_i, s_{-i}) \leq c_i(s'_i, s_{-i}).$$

It is straightforward to associate with each game that uses cost functions a customary strategic game by using

$$p_i(s) := -c_i(s).$$

We now define a ***congestion game*** for n players as follows. Assume a non-empty finite set E of ***facilities***, for example road segments. Given a

player i his set of strategies is a set of non-empty subsets of E, i.e. $S_i \subseteq \mathcal{P}(E) \setminus \{\{\emptyset\}\}$.

We define the cost functions c_i as follows. First, we introduce the ***delay function*** $d_j : \{1, \ldots, n\} \to \mathbb{R}$ for using facility $j \in E$; $d_j(k)$ is the ***delay*** for using facility j when there are k users of j. Next, we define a function $u_j : S_1 \times \cdots \times S_n \to \{1, \ldots, n\}$ by

$$u_j(s) := |\{r \in \{1, \ldots, n\} \mid j \in s_r\}|.$$

So $u_j(s)$ is the number of users of facility j given the joint strategy s. Finally, we define the cost function by

$$c_i(s) := \sum_{j \in s_i} d_j(u_j(s)).$$

So $c_i(s)$ is the aggregate delay incurred by player i when each player j selected the set of facilities s_j.

The following important result clarifies our interest in the congestion games.

Theorem 18 (Congestion) *Every congestion game is a potential game.*

Proof. We define

$$P(s) := \sum_{j \in s_1 \cup \ldots \cup s_n} \sum_{k=1}^{u_j(s)} d_j(k),$$

Intuitively, $P(s)$ is the sum of accumulated delays for each facility. We prove that P is indeed a potential.

First, we extend each function d_j to $\{0, 1, \ldots, n\}$ by putting $d_j(0) := 0$. We have then

$$P(s) = \sum_{j \in E} \sum_{k=0}^{u_j(s)} d_j(k), \tag{5.1}$$

since for $j \in E \setminus (s_1 \cup \ldots \cup s_n)$ we have $u_j(s) = 0$.

Recall that χ_A denotes the set characteristic function for the set A, i.e., $\chi_(A)(j) = 1$ if $j \in A$ and 0 otherwise. We have then for $i \in \{1, \ldots, n\}$

$$c_i(s) = \sum_{j \in E} d_i(u_j(s)) \chi_{s_i}(j). \tag{5.2}$$

From (5.1) and (5.2) it follows that for $i \in \{1, \ldots, n\}$

$$P(s) - c_i(s) = \sum_{j \in E} \sum_{k=0}^{u_j(s) - \chi_{s_i}(j)} d_j(k) \tag{5.3}$$

and

$$P(s'_i, s_{-i}) - c_i(s'_i, s_{-i}) = \sum_{j \in E} \sum_{k=0}^{u_j(s'_i, s_{-i}) - \chi_{s'_i}(j)} d_j(k). \tag{5.4}$$

But for $j \in E$ we have $u_j(s) - \chi_{s_i}(j) = u_j(s'_i, s_{-i}) - \chi_{s'_i}(j)$, so from (5.3) and (5.4) it follows that P is a potential. □

Example 13 We now discuss the so-called ***Braess paradox*** showing that adding new roads to a road network can lead to an increased travel time. To discuss it we use the game theoretic concepts of a Nash equilibrium, strictly dominant strategies and a social welfare. Consider the road network given in Figure 5.4.

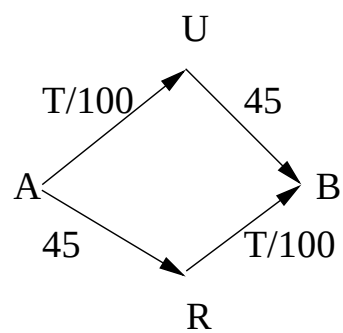

Figure 5.4: A road network

Assume that there are 4000 players (drivers), travelling from A to B. Each of them has two strategies consisting of a road A - U - B or A - R - B. The delays for each facility (road segment) are indicated in the figure. So if T drivers choose the road segment A - U (or R - B), then the delay is T/100. The delay for the other two road segments is constant.

It is easy to see that a joint strategy is a Nash equilibrium iff the drivers evenly split among the two possible roads, that is 2000 players choose one strategy and 2000 the other strategy. The resulting cost (travel time) for each player (driver) equals 2000/100 + 45 = 45 + 2000/100 = 65.

Suppose now that a new, fast, road from U to R is added to the network with delay 0, see Figure 5.5.

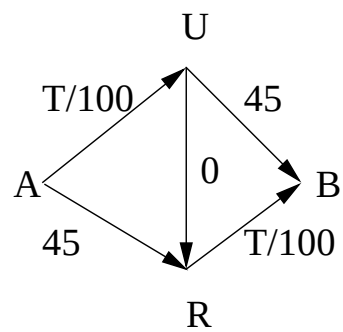

Figure 5.5: An augmented road network

Now each player (driver) has three possible strategies (routes): A - U - B, A - R - B, and A - U - R - B. It is easy to see that in this new congestion game for each player A - U - R - B is a strictly dominant strategy. Consequently, by the Strict Dominance Note 1 this new game has a unique Nash equilibrium that consists of these strictly dominant strategies. Moreover, by the Best Response Dynamics Note 14, all executions of the best response dynamics terminate in this unique Nash equilibrium.

Now, the resulting travel time for each driver equals 4000/100 + 4000/100 = 80, so it increased. This shows that adding the new road segment, in this case U - R, can result in a longer travel time. It is easy to check that this paradox remains in force as long as the delay for using U - R is smaller than 5. □

Exercise 13 Prove that in the above example for each player A - U - R - B is indeed a strictly dominant strategy. □

A special case of congestion games are the ***fair cost sharing games***. In these games each facility $j \in E$ has a cost $c_j \in \mathbb{R}$ associated with it. Then the delay function for a facility is obtained by dividing its cost equally between the users. So we use

$$d_j(u_j(s)) := \frac{c_j}{u_j(s)}$$

in the definition of the congestion game. Consequently

$$c_i(s) := \sum_{j \in s_i} \frac{c_j}{u_j(s)}.$$

Fair cost sharing games form a natural class of congestion games in which the costs decrease when the number of users of the shared facilities increases. In this context the delay function should be viewed as the charge for the use of the facility.