Potential functions and Nash equilibria

Elizaveta Rebrova

ORF387, September 2022

The best response dynamic is a process that can find an equilibrium strategy. It starts from any joint strategy $(s_1, \dots, s_n)$, where s_i is the strategy employed by player i. If there is an s_i so that it is not the best response to all other s_j's, in the next step we consider a joint strategy $(s_1, \dots, s_i', \dots, s_n)$, where s_i' is the best response (and the other s_j did not change). Repeat the process until we cannot find a player with the best response strategy different from the one they use. If the process ends, this must be a Nash equilibrium (as no player has an incentive to change their strategy).

This process might never end but keep going in an infinite loop, as in the *mixed pennies game* example. Recall the game: Player 1 and Player 2 choose H or T. Player 1 gets payoff 1 if the joint strategy is (T, T) or (H, H), and -1 if (T, H) or (H, T) is played. Player 2 gets a payoff negative of what Player 1 gets (one player pays a penny to the other). The best response dynamic starting from the joint strategy (H, T) looks like

- Player 1's best response changes it to the joint strategy (T, T)
- Then, Player 2's best response changes it to (T, H).
- Then, Player 1's best response changes it to the joint strategy (H, H).
- Then, Player 2's best response changes it to the joint strategy (H, T), and we are back to the joint strategy considered first, it can be improved by Player 2 by changing it...

It is no surprise the process does not end as there is no Nash equilibrium in the matching pennies game. However, even when there is an equilibrium, we might end up in an infinite loop starting with some joint strategies.

The potential function is a single function that expresses the incentive of all players to change their strategy. Formally,

Definition 1. *A function Φ from the set of all joint strategies to $\mathbb{R}$ is a **potential function** if for any 2 joint strategies $s = (s_1, \dots, s_i, \dots, s_n)$ and $s' = (s_1, \dots, s_i', \dots, s_n)$ such that*

- *they differ only at one position i (so, the player i can change strategy s to strategy s')*

- *the payoff of player i if the strategy s' was played is higher than the payoff of i if s was played (so, there is an incentive for the player i to do the change from s to s'),*

$$\Phi(s') < \Phi(s).$$

The key observation is that if a finite game has a potential function, then it has a pure Nash equilibrium and the best response dynamic must converge it.

Why? Indeed, every best response decreases the potential function, so every step of the best response dynamic gives a new strategy: all the joint strategies we've seen earlier had higher values of Φ! Since there are finitely many pure strategies in the game, the process must stop, at least after we look at all the strategies and find the one with the minimal value of the potential (it can also terminate earlier, not at the global minimum of Φ).

- Unfortunately, potential functions can be hard to find even when they exist
- But if someone gives you a function Φ and you can show that its value must decay if any player changes their strategy to increases their payoff, then you can (a) claim that pure Nash equilibrium exists and (b) run the best response dynamic starting from any joint strategy to find it.
- Mixed Nash equilibrium always exist for finite games (Nash theorem). The proof of this fact is nontrivial and is roughly based on the generalization of the better response dynamic for mixed strategies. One can prove that it always converges to an equilibrium point (using Brower's fixed point theorem).

Examples of potential functions: **1. Prisoner's dillemma.**

		Suspect 2: NC	Suspect 2: C
Suspect 1	NC	$-1, -1$	$-10, 0$
	C	$0, -10$	$-4, -4$

Figure 6.2: Prisoner's Dilemma

A potential function = *the number of players who selected the "not confess" strategy.* Why? Here, the best response for every player in every situation is to confess (it is a dominating strategy). So, swithching to the best response strategy is always changing from (NC) to (C). Clearly, this change increases the number of players who decided to confess and decreases the potential function.

2. NBA teams choosing cities. The executives of n NBA teams are trying to choose where to locate their teams, and they have k different city choices (for example, LA, SF, NY, etc), denoted by $1, 2, \ldots, k$.[1] The profit (determined by the number of local basketball fans) of city j is F_j. If city j is selected by t teams, then they split the profits evenly, each getting F_j/t. Is there an equilibium city choice for all the teams?

In this case, a potential function Φ of a certain strategy (assignment of the teams to certain cities) can be defined as

$$-\sum_{j=1}^{k}\sum_{t=1}^{n_j}\frac{F_j}{t}, \quad \text{where } n_j \text{ denote the number of teams that select city } j.$$

For example, if two teams chose city 1, four teams chose city 2, city 3 has no teams and city 4 has 1 team, this function is equal to

$$-F_1 - \frac{F_1}{2} - F_2 - \frac{F_2}{2} - \frac{F_2}{3} - \frac{F_2}{4} - F_4.$$

In general, if a team gets better payoff from moving from city i to city j, then, its current profit F_i/n_i is less than its prospective profit $F_j/(n_j+1)$. At the same time, the new strategy results in the value of the function Φ that is smaller by exactly $F_j/(n_j+1) - F_i/n_i > 0$ (check this!). So, Φ is indeed a potential function and there is always an equilibrium choice of the cities.

[1] While in essentially all other countries of the world it is pretty rare for sports teams to relocate, this is pretty common in the US: `https://en.wikipedia.org/wiki/List_of_relocated_National_Basketball_Association_teams`. Thanks to Micklos Racz for this example.