

Lecture 10

Lecturer: Aaron Roth

Scribe: Aaron Roth

The Price of Anarchy and Stability

Up until now, we have been focussing on *how* agents playing together in a game, in a decentralized manner, might arrive at an equilibrium. For different equilibrium concepts, and in different settings, we have seen different plausible ways in which this might happen (i.e. best response dynamics in congestion games, no-regret dynamics in zero sum games, and no-swap-regret dynamics in general games). But suppose players *do* reach an equilibrium. What then? What can we say about the quality of the outcome that has been reached?

This is where the *price of anarchy* and *price of stability* come in. They measure how bad things *can* and *must* get respectively, when players play according to an equilibrium. In this lecture we will study this concept with respect to the *social cost* objective and *Nash* equilibrium, but more generally, it makes sense to study the price of anarchy and stability for any objective of interest, over any class of equilibria.

In order to talk about the quality of a game state, we must define what our objective function is. We will think about games in which players have individual cost functions $c_i : A \rightarrow \mathbb{R}$.

Let $\text{Objective} : A \rightarrow \mathbb{R}$ measure the cost of game states a . We will generally be interested in the social cost objective which measures the sum cost of all of the players:

$$\text{Objective}(a) = \sum_{i=1}^n c_i(a)$$

but we could also study other objectives. Note that *smaller* objective values are better.

We define OPT to be the optimal value the objective function ever takes on any action profile. This is the quality of the solution we could obtain if we had dictatorial control, and could mandate the action that everyone took:

$$\text{OPT} = \min_{a \in A} \text{Objective}(a)$$

On the other hand, in a game, players can make decisions independently, and we are interested in how much worse things can be in rational solutions. The price of anarchy measures how bad the objective can be (compared to OPT) in the worst case, if we assume nothing other than that players play according to some Nash equilibrium.

Definition 1 *The price of anarchy of a game G is:*

$$\text{PoA} = \max_{a: a \text{ is a Nash equilibrium of } G} \frac{\text{Objective}(a)}{\text{OPT}}$$

i.e. it is how much worse things can be in the worst case.

In contrast, the price of stability measures how much worst things *must be* (compared to OPT) if we assume that players are playing according to a Nash equilibrium – it measures the quality of the best Nash equilibrium:

Definition 2 *The price of stability of a game G is:*

$$\text{PoS} = \min_{a: a \text{ is a Nash equilibrium of } G} \frac{\text{Objective}(a)}{\text{OPT}}$$

The names are appropriate. The price of anarchy measures how bad things can get if we let everyone act for themselves – if we have anarchy – and assume only that they are rational enough to reach equilibrium. In contrast, if we have the power to suggest to players how they should play, we could suggest that they play the *best* Nash equilibrium. But if we want our suggestions to be stable, we must

suggest some stable state (i.e. an equilibrium). The price of stability tells us how bad things must be, even if we get to pick our favorite stable state.

Note that we have defined the price of anarchy and stability with respect to Nash equilibrium, but we could equally well define them with respect to any other solution concept. Recalling the ones we have studied: pure strategy Nash equilibrium (PSNE), mixed strategy Nash equilibrium (MSNE), Correlated Equilibrium (CE), and Coarse Correlated Equilibrium (CCE), we could define $PoA(PSNE)$, $PoA(MSNE)$, $PoA(CE)$, and $PoA(CCE)$. Note that if we did this, for any fixed game, we would have:

$$PoA(PSNE) \leq PoA(MSNE) \leq PoA(CE) \leq PoA(CCE)$$

(can you see why?)

For this lecture, we will restrict our attention to the price of anarchy defined over Nash equilibria however.

Recall the fair cost sharing game that we discussed when we covered congestion games: It is an n player m facility congestion game in which each facility j has some weight w_j and we have:

$$\ell_j(k) = \frac{w_j}{k} \quad c_i(a) = \sum_{j \in a_i} \ell_j(n_j(a))$$

i.e. all agents playing on a resource j uniformly split the cost w_j of building the resource (the more people using it the cheaper it is), and the total cost of an agent is the sum over all of his resource costs. Note that for this game, the social cost objective:

$$\text{Objective}(a) = \sum_{i=1}^n c_i(a) = \sum_{j \in a_1 \cup \dots \cup a_n} w_j$$

is exactly equal to the total cost of the resources built.

We saw in class an instance proving the following theorem:

Theorem 3 *For fair cost sharing games:*

$$PoS(PSNE) \geq H_n = \Omega(\log n)$$

where $H_n = \sum_{i=1}^n 1/i$ is the n 'th harmonic number.

Proof We gave an example in class in which the only Nash equilibrium had objective value H_n , but the optimal solution had objective value $1 + \epsilon$ for arbitrarily small ϵ . ■

We can also show a matching upper bound on the price of stability.

Theorem 4 *For fair cost sharing games:*

$$PoS(PSNE) \leq H_n = O(\log n)$$

Proof We recall that congestion games have an exact potential function:

$$\phi(a) = \sum_{j: n_j(a) \geq 1} \sum_{k=1}^{n_j(a)} \ell_j(k)$$

and that when players play best response moves, the potential function ϕ only decreases. In this case, we have:

$$\begin{aligned}
\phi(a) &= \sum_{j:n_j(a) \geq 1} \sum_{k=1}^{n_j(a)} \frac{w_j}{k} \\
&= \sum_{j \in a_1 \cup \dots \cup a_n} w_j \cdot \sum_{k=1}^{n_j(a)} \frac{1}{k} \\
&\leq \sum_{j \in a_1 \cup \dots \cup a_n} w_j \cdot H_n \\
&= H_n \cdot \text{Objective}(a)
\end{aligned}$$

Also note, that clearly:

$$\text{Objective}(a) \leq \phi(a)$$

Let a^* be a state such that $\text{Objective}(a^*) = \text{OPT}$. Imagine starting at state a^* and then running best response dynamics until it converges to a pure strategy Nash equilibrium a' (as we know it must in a congestion game). Since best response dynamics only decreases the potential function, we know:

$$\begin{aligned}
\text{Objective}(a') &\leq \phi(a') \\
&\leq \phi(a^*) \\
&\leq H_n \text{Objective}(a^*) \\
&= H_n \cdot \text{OPT}
\end{aligned}$$

which proves the claim. ■

Together, we have exactly pinned down the price of stability for fair cost sharing games. We can also study the price of anarchy, which is much worse:

Theorem 5 *In fair cost sharing games:*

$$PoA(PSNE) \geq n$$

Proof Consider a game with two facilities $F = \{1, 2\}$ and $A_i = \{\{1\}, \{2\}\}$ for all players i . The cost of facility 1 is $w_1 = (1 + \epsilon)$, and the cost of facility 2 is $w_2 = n$. Consider the action profile a in which $a_i = \{2\}$ for all players i . Then the cost of every player is $w_2/n = 1$, and the social cost is $\text{Objective}(a) = n$. Moreover, a is a Nash equilibrium, since if any player deviates to $a_i = \{1\}$, he will experience cost $1 + \epsilon > 1$. However, the optimal solution a^* sets $a_i^* = \{1\}$ for every player i , and this has social cost $\text{OPT} = 1 + \epsilon$. Since we can set ϵ as small as we like, this proves the claim. ■

Again, we can show that this bound is tight:

Theorem 6 *In fair cost sharing games:*

$$PoA(PSNE) \leq n$$

Proof Let a^* be an action profile such that $\text{Objective}(a^*) = \text{OPT}$. We claim that for every pure strategy Nash equilibrium a :

$$c_i(a) \leq n \cdot c_i(a^*)$$

because by the Nash equilibrium condition, for every player i :

$$\begin{aligned}c_i(a) &\leq c_i(a_i^*, a_{-i}) \\&\leq \sum_{j \in a_i^*} \ell_j(\max(n_j(a), 1)) \\&= \sum_{j \in a_i^*} \frac{w_j}{\max(n_j(a), 1)} \\&\leq \sum_{j \in a_i^*} w_j \\&= n \cdot \sum_{j \in a_i^*} \frac{w_j}{n} \\&\leq n \cdot c_i(a^*)\end{aligned}$$

Since this holds term by term, certainly $\sum_{i=1}^n c_i(a) \leq n \sum_{i=1}^n c_i(a^*)$ which is what we wanted to show.

■