The Price of Anarchy and Stability

Aaron Roth

University of Pennsylvania

February 20 2025

Overview

- So far we have focused on *how* agents playing together in a game might arrive at an equilibrium.

Overview

- So far we have focused on *how* agents playing together in a game might arrive at an equilibrium.
- For different equilibrium concepts, in different settings, we have seen different plausible ways in which this might happen.

Overview

- So far we have focused on *how* agents playing together in a game might arrive at an equilibrium.
- For different equilibrium concepts, in different settings, we have seen different plausible ways in which this might happen.
- But suppose players *do* reach an equilibrium. What then?

Overview

- So far we have focused on *how* agents playing together in a game might arrive at an equilibrium.
- For different equilibrium concepts, in different settings, we have seen different plausible ways in which this might happen.
- But suppose players *do* reach an equilibrium. What then?
- What can we say about the quality of the outcome that has been reached?

Overview

- So far we have focused on *how* agents playing together in a game might arrive at an equilibrium.
- For different equilibrium concepts, in different settings, we have seen different plausible ways in which this might happen.
- 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

Overview

- So far we have focused on *how* agents playing together in a game might arrive at an equilibrium.
- For different equilibrium concepts, in different settings, we have seen different plausible ways in which this might happen.
- 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
- We'll study this question for Nash equilibria, but more generally its sensible to study for any of the equilibrium concepts we have seen.

The Designer's Goal

1. In order to talk about the quality of a game state, we must define what our objective function is.

The Designer's Goal

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

The Designer's Goal

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

The Designer's Goal

1. In order to talk about the quality of a game state, we must define what our objective function is.
2. We will think about games in which players have individual cost functions $c_i : A \to \mathbb{R}$.
3. Let $\mathrm{Objective} : A \to \mathbb{R}$ measure the cost of game states a.
4. We will generally be interested in the social cost objective: the sum cost of all of the players:

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

The Designer's Goal

1. In order to talk about the quality of a game state, we must define what our objective function is.
2. We will think about games in which players have individual cost functions $c_i : A \to \mathbb{R}$.
3. Let $\mathrm{Objective} : A \to \mathbb{R}$ measure the cost of game states a.
4. We will generally be interested in the social cost objective: the sum cost of all of the players:

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

5. More generally we could be interested in other things. Note in this case, smaller values are better.

The Designer's Goal

1. Define OPT to be the optimal value the objective function takes on any action profile. This is the quality of the solution we could obtain if we had dictatorial control:

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

The Designer's Goal

1. Define OPT to be the optimal value the objective function takes on any action profile. This is the quality of the solution we could obtain if we had dictatorial control:

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

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

The Designer's Goal

1. Define OPT to be the optimal value the objective function takes on any action profile. This is the quality of the solution we could obtain if we had dictatorial control:

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

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

Definition

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

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

An Optimistic Perspective

1. The price of anarchy pessimistically measures how much things can go wrong if we might end up in an *arbitrary* Nash equilibrium.

An Optimistic Perspective

1. The price of anarchy pessimistically measures how much things can go wrong if we might end up in an *arbitrary* Nash equilibrium.
2. What if we get to choose the (equilibrium) outcome — how bad *must* things get?

Definition

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

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

An Optimistic Perspective

1. The price of anarchy pessimistically measures how much things can go wrong if we might end up in an *arbitrary* Nash equilibrium.
2. What if we get to choose the (equilibrium) outcome — how bad *must* things get?

Definition

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

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

3. The names are appropriate/evocative.

An Optimistic Perspective

1. The price of anarchy pessimistically measures how much things can go wrong if we might end up in an *arbitrary* Nash equilibrium.
2. What if we get to choose the (equilibrium) outcome — how bad *must* things get?

Definition

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

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

3. The names are appropriate/evocative.
4. We have defined Price of Anarchy (POA) and Price of Stability (PoS) for Nash equilibria, but we could have defined them for any of our equilibrium concepts. Observe:

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

(why?)

Fair Cost Sharing Games

1. This lecture: We'll restrict attention to pure strategy Nash equilibria.

Fair Cost Sharing Games

1. This lecture: We'll restrict attention to pure strategy Nash equilibria.
2. Recall the fair cost sharing game (a congestion game): An n player m facility congestion game in which each facility j has some weight w_j and we have:

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

Fair Cost Sharing Games

1. This lecture: We'll restrict attention to pure strategy Nash equilibria.
2. Recall the fair cost sharing game (a congestion game): An n player m facility congestion game in which each facility j has some weight w_j and we have:

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

3. i.e. all agents playing on a resource j uniformly split the cost w_j of building the resource, and the total cost of an agent is the sum over all of his resource costs.

Fair Cost Sharing Games

1. This lecture: We'll restrict attention to pure strategy Nash equilibria.
2. Recall the fair cost sharing game (a congestion game): An n player m facility congestion game in which each facility j has some weight w_j and we have:

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

3. i.e. all agents playing on a resource j uniformly split the cost w_j of building the resource, and the total cost of an agent is the sum over all of his resource costs.
4. The social cost in this case is the total cost of resources built:

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

Fair Cost Sharing Games

Theorem

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.*

Fair Cost Sharing Games

Theorem

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.*

To prove a lower bound, we only need to give an example...

Fair Cost Sharing Games

Fair Cost Sharing Games

Theorem

For fair cost sharing games:

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

Fair Cost Sharing Games

Theorem

For fair cost sharing games:

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

To prove an upper bound, we need a more sophisticated argument because we need to show something for *all* such games.

Proof

1. 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 it decreases with best response moves.

Proof

1. 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 it decreases with best response moves.

2. We can compute:

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

Proof

1. 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 it decreases with best response moves.

2. We can compute:

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

Proof

1. 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 it decreases with best response moves.

2. We can compute:

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

Proof

1. 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 it decreases with best response moves.

2. We can compute:

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

Proof

1. Also observe:

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

Proof

1. Also observe:
$$\mathrm{Objective}(a) \leq \phi(a)$$
2. Thus:
$$\mathrm{Objective}(a) \leq \phi(a) \leq H_n \cdot \mathrm{Objective}(a)$$

Proof

1. Also observe:
$$\mathrm{Objective}(a) \le \phi(a)$$
2. Thus:
$$\mathrm{Objective}(a) \le \phi(a) \le H_n \cdot \mathrm{Objective}(a)$$
3. So lets conduct a thought experiment...

Proof

1. Also observe:
$$\mathrm{Objective}(a) \leq \phi(a)$$
2. Thus:
$$\mathrm{Objective}(a) \leq \phi(a) \leq H_n \cdot \mathrm{Objective}(a)$$
3. So lets conduct a thought experiment...
4. Let a^* be a state such that $\mathrm{Objective}(a^*) = \mathrm{OPT}$.

A Thought Experiment

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

1. Imagine starting at state a^* and then running best response dynamics until it converges to a PSNE a'.

A Thought Experiment

$$\mathrm{Objective}(a) \leq \phi(a) \leq H_n \cdot \mathrm{Objective}(a)$$

1. Imagine starting at state a^* and then running best response dynamics until it converges to a PSNE a'.
2. We know:

$$\mathrm{Objective}(a') \quad \leq \quad \phi(a')$$

A Thought Experiment

$$\text{Objective}(a) \le \phi(a) \le H_n \cdot \text{Objective}(a)$$

1. Imagine starting at state a^* and then running best response dynamics until it converges to a PSNE a'.
2. We know:

$$\begin{aligned} \text{Objective}(a') &\le \phi(a') \\ &\le \phi(a^*) \end{aligned}$$

A Thought Experiment

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

1. Imagine starting at state a^* and then running best response dynamics until it converges to a PSNE a'.
2. We know:

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

A Thought Experiment

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

1. Imagine starting at state a^* and then running best response dynamics until it converges to a PSNE a'.
2. We know:

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

A Thought Experiment

$$\mathrm{Objective}(a) \leq \phi(a) \leq H_n \cdot \mathrm{Objective}(a)$$

1. Imagine starting at state a^* and then running best response dynamics until it converges to a PSNE a'.
2. We know:

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

3. Tada!

The Price of Anarchy

The price of anarchy can only be worse, and it is...

Theorem

In fair cost sharing games:

$$PoA(PSNE) \geq n$$

The Price of Anarchy

The price of anarchy can only be worse, and it is...

Theorem

In fair cost sharing games:

$$PoA(PSNE) \geq n$$

Once again, to prove a lower bound we just need an example...

The Price of Anarchy

The Price of Anarchy

Theorem

In fair cost sharing games:

$$PoA(PSNE) \leq n$$

The Price of Anarchy

Theorem

In fair cost sharing games:

$$PoA(PSNE) \leq n$$

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

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

Why?

Proof

By the Nash equilibrium condition, for every player i:

$$c_i(a) \quad \leq \quad c_i(a_i^*, a_{-i})$$

Proof

By the Nash equilibrium condition, for every player i:

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

Proof

By the Nash equilibrium condition, for every player i:

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

Proof

By the Nash equilibrium condition, for every player i:

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

Proof

By the Nash equilibrium condition, for every player i:

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

Proof

By the Nash equilibrium condition, for every player i:

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

Proof

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: $\sum_{i=1}^{n} c_i(a) \leq n \sum_{i=1}^{n} c_i(a^*)$.

Thanks!

See you next class — stay healthy!