Truthful, Pareto Optimal Exchange Without Money

Aaron Roth

University of Pennsylvania

February 25, 2025

Overview

- Up until now we have studied the behaviour of individuals in already defined games.

Overview

- Up until now we have studied the behaviour of individuals in already defined games.
- This will be the first lecture on "Mechanism Design"

Overview

- Up until now we have studied the behaviour of individuals in already defined games.
- This will be the first lecture on “Mechanism Design”
- Designing the rules of the game to achieve our goals.

Overview

- Up until now we have studied the behaviour of individuals in already defined games.
- This will be the first lecture on "Mechanism Design"
- Designing the rules of the game to achieve our goals.
- We'll begin our study with the classical "House Allocation Problem" by Shapley and Scarf.

Overview

- Up until now we have studied the behaviour of individuals in already defined games.
- This will be the first lecture on "Mechanism Design"
- Designing the rules of the game to achieve our goals.
- We'll begin our study with the classical "House Allocation Problem" by Shapley and Scarf.
- And study the Top Trading Cycles Algorithm (attributed to David Gale).

“House Allocation”

1. Each individual comes to market with a single good (a “house”), but could be e.g. a kidney.

“House Allocation”

1. Each individual comes to market with a single good (a “house”), but could be e.g. a kidney.
2. Each individual has a strict preference ordering over other goods in the market.

"House Allocation"

1. Each individual comes to market with a single good (a "house"), but could be e.g. a kidney.
2. Each individual has a strict preference ordering over other goods in the market.
3. The question: How can we both:

“House Allocation”

1. Each individual comes to market with a single good (a “house”), but could be e.g. a kidney.
2. Each individual has a strict preference ordering over other goods in the market.
3. The question: How can we both:
   3.1 Coordinate an exchange to arrive at a *good* allocation, and

“House Allocation”

1. Each individual comes to market with a single good (a “house”), but could be e.g. a kidney.
2. Each individual has a strict preference ordering over other goods in the market.
3. The question: How can we both:
    3.1 Coordinate an exchange to arrive at a *good* allocation, and
    3.2 Do so in a way such that it is a dominant strategy for everyone to report their true preferences.

“House Allocation”

1. Each individual comes to market with a single good (a “house”), but could be e.g. a kidney.
2. Each individual has a strict preference ordering over other goods in the market.
3. The question: How can we both:
   3.1 Coordinate an exchange to arrive at a *good* allocation, and
   3.2 Do so in a way such that it is a dominant strategy for everyone to report their true preferences.
4. Doing both is important. If we merely guarantee a “good” allocation, we only know it is “good” w.r.t. reported preferences. But it might be bad w.r.t. real preferences!

"House Allocation"

1. Each individual comes to market with a single good (a "house"), but could be e.g. a kidney.
2. Each individual has a strict preference ordering over other goods in the market.
3. The question: How can we both:
   3.1 Coordinate an exchange to arrive at a *good* allocation, and
   3.2 Do so in a way such that it is a dominant strategy for everyone to report their true preferences.
4. Doing both is important. If we merely guarantee a "good" allocation, we only know it is "good" w.r.t. reported preferences. But it might be bad w.r.t. real preferences!
5. Houses are a toy example. Kidney exchange is a real one (needs a solution without money).

A Model

1. There are n agents $i \in P$ who each come to market with a good h_i.
2. Each agent has a strict preference ordering $\succ_i$ over all of the goods $h_1, \ldots, h_n$. (i.e. for every pair j, k either $h_j \succ_i h_k$ or $h_k \succ_i h_j$, and this ordering is transitive – so each agent just has a rank order list of goods. In particular, this ranking includes an agents own good h_i.

A Model

1. There are n agents $i \in P$ who each come to market with a good h_i.
2. Each agent has a strict preference ordering $\succ_i$ over all of the goods $h_1, \ldots, h_n$. (i.e. for every pair j, k either $h_j \succ_i h_k$ or $h_k \succ_i h_j$, and this ordering is transitive – so each agent just has a rank order list of goods. In particular, this ranking includes an agents own good h_i.

We wish to design an algorithm which will induce a game played by the players. The algorithm will take as input the reported preferences $\succ_i$ of each player, and output a permutation μ of the goods. This induces a game: the strategy space for each player is the set of preference orderings $\succ_i$, the utility function is defined by their true preferences.

What is Good?

What is Good?

Definition

An allocation μ is *Pareto sub-optimal* if there exists an allocation ν such that for every i:

$$\nu(i) \succeq_i \mu(i)$$

and for some j;

$$\nu(j) \succ_j \mu(j)$$

i.e. everybody is at least as happy with their allocation in ν, and at least one person is strictly happier. In this case, we say that ν *Pareto-dominates* μ.

If μ is not Pareto sub-optimal, then it is *Pareto optimal*.

What is Good?

Definition

An allocation μ is *Pareto sub-optimal* if there exists an allocation ν such that for every i:

$$\nu(i) \succeq_i \mu(i)$$

and for some j;

$$\nu(j) \succ_j \mu(j)$$

i.e. everybody is at least as happy with their allocation in ν, and at least one person is strictly happier. In this case, we say that ν *Pareto-dominates* μ.
If μ is not Pareto sub-optimal, then it is *Pareto optimal*.

It should not be possible to simultaneously improve for everyone.

What about Incentives?

Definition

A is individually rational if for every player i, every preference vector $\succ_i$, and every set of reports of the other players $\succ_{-i}$, if $\mu = A(\succ_i, \succ_{-i})$ then:

$$\mu(i) \succeq_i h_i$$

What about Incentives?

Definition

A is individually rational if for every player i, every preference vector $\succ_i$, and every set of reports of the other players $\succ_{-i}$, if $\mu = A(\succ_i, \succ_{-i})$ then:

$$\mu(i) \succeq_i h_i$$

People should not be harmed by participating... A minimal goal; we want more.

What about Incentives?

Definition

A is individually rational if for every player i, every preference vector $\succ_i$, and every set of reports of the other players $\succ_{-i}$, if $\mu = A(\succ_i, \succ_{-i})$ then:

$$\mu(i) \succeq_i h_i$$

People should not be harmed by participating... A minimal goal; we want more.

Definition

A mechanism A is dominant-strategy incentive compatible if it is a dominant strategy for everyone to report their true preferences. i.e. if for all $\succ_i, \succ_{-i}, \succ'_i$, if

$$\mu = A(\succ_i, \succ_{-i}) \quad \text{and } \nu = A(\succ'_i, \succ_{-i})$$

then $\mu(i) \succeq_i \nu(i)$

Top Trading Cycles

Algorithm 1 The top trading cycles algorithm

TTC$(\succ_1, \ldots, \succ_n)$

Let $S_1 = P$ be the set of all agents. Set a counter $t = 1$.

while $|S_i| > 0$ **do**

Construct a graph $G_t = (V_t, E_t)$ where $V_t = S_t$ and for each $i, j \in V_t$, the directed edge $(i, j) \in E_t$ if and only if $h_j \succ_i h_k$ for all other $k \in V_t$. i.e. this is the graph that results when every agent "points to" their favorite remaining good.

Find any cycle C_t in G_t and clear all trades along it: i.e. for every directed edge $(i, j) \in C_t$ set $\mu(i) = j$.

Set $S_{t+1} = S_t$ and remove all cleared agents: for each i : $(i, j) \in C_t$, set $S_{t+1} \leftarrow S_{t+1} - \{i\}$. Increment t $(t \leftarrow t + 1)$.

end while

Output μ.

Analysis

1. First: establish the algorithm halts at all.

Analysis

1. First: establish the algorithm halts at all.
2. Enough to show we find a cycle at every round, since this removes at least one agent: Convergence after $\leq n$ rounds.

Analysis

1. First: establish the algorithm halts at all.
2. Enough to show we find a cycle at every round, since this removes at least one agent: Convergence after $\leq n$ rounds.
3. Indeed:

Lemma

In each graph G_t constructed by the algorithm, there is at least one cycle C_t, and every agent is part of at most one cycle.

Analysis

1. First: establish the algorithm halts at all.
2. Enough to show we find a cycle at every round, since this removes at least one agent: Convergence after $\leq n$ rounds.
3. Indeed:

Lemma

In each graph G_t constructed by the algorithm, there is at least one cycle C_t, and every agent is part of at most one cycle.

4. **Proof**: by construction, G_t is a directed graph in which every vertex has out-degree exactly one. (So by starting at any vertex and following edges forward, we must find a cycle).

Interlude: Example

5 agents:

$\succ_1$: $2 \succ 5 \succ 3 \succ 1 \succ 4$

$\succ_2$: $3 \succ 1 \succ 5 \succ 4 \succ 2$

$\succ_3$: $1 \succ 2 \succ 3 \succ 4 \succ 5$

$\succ_4$: $1 \succ 3 \succ 5 \succ 4 \succ 2$

$\succ_5$: $4 \succ 1 \succ 3 \succ 2 \succ 5$

Interlude: Example

5 agents:

$\succ_1$: $2 \succ 5 \succ 3 \succ 1 \succ 4$

$\succ_2$: $3 \succ 1 \succ 5 \succ 4 \succ 2$

$\succ_3$: $1 \succ 2 \succ 3 \succ 4 \succ 5$

$\succ_4$: $1 \succ 3 \succ 5 \succ 4 \succ 2$

$\succ_5$: $4 \succ 1 \succ 3 \succ 2 \succ 5$

$$\mu(1) = 2, \mu(2) = 3, \mu(3) = 1, \mu(4) = 5, \mu(5) = 4$$

Analysis

Theorem

The Top Trading Cycles algorithm produces a Pareto optimal allocation μ *on every input* $\succ$.

Proof

1. Suppose not. So there is some allocation ν that Pareto dominates μ. What does ν look like?

Proof

1. Suppose not. So there is some allocation ν that Pareto dominates μ. What does ν look like?
2. Observe: every agent TTC cleared in cycle C_1 must receive an identical allocation in ν. Why?

Proof

1. Suppose not. So there is some allocation ν that Pareto dominates μ. What does ν look like?
2. Observe: every agent TTC cleared in cycle C_1 must receive an identical allocation in ν. Why?
3. Next: every agent TTC cleared in cycle C_2 must receive an identical allocation in ν: since these agents are receiving their first choice good from the set $P - C_1$ in μ, and $\nu(i) = \mu(i)$ for every $i \in C_1$

Proof

1. Suppose not. So there is some allocation ν that Pareto dominates μ. What does ν look like?
2. Observe: every agent TTC cleared in cycle C_1 must receive an identical allocation in ν. Why?
3. Next: every agent TTC cleared in cycle C_2 must receive an identical allocation in ν: since these agents are receiving their first choice good from the set $P - C_1$ in μ, and $\nu(i) = \mu(i)$ for every $i \in C_1$
4. Inductively, if $\nu(i) = \mu(i)$ for every $i \in C_1 \cup \ldots \cup C_k$ for $k \leq t$, then We must also have that $\nu(i) = \mu(i)$ for every $i \in C_{t+1}$.

Proof

1. Suppose not. So there is some allocation ν that Pareto dominates μ. What does ν look like?
2. Observe: every agent TTC cleared in cycle C_1 must receive an identical allocation in ν. Why?
3. Next: every agent TTC cleared in cycle C_2 must receive an identical allocation in ν: since these agents are receiving their first choice good from the set $P - C_1$ in μ, and $\nu(i) = \mu(i)$ for every $i \in C_1$
4. Inductively, if $\nu(i) = \mu(i)$ for every $i \in C_1 \cup \ldots \cup C_k$ for $k \leq t$, then We must also have that $\nu(i) = \mu(i)$ for every $i \in C_{t+1}$.
5. Continuing through $t = n$, we have that $\mu = \nu$, a contradiction.

Analysis

Theorem

The Top Trading Cycles algorithm is individually rational.

Analysis

Theorem

The Top Trading Cycles algorithm is individually rational.

Why?

Analysis

Theorem

The Top Trading Cycles Algorithm is Dominant Strategy Incentive Compatible.

Proof (Sketch)

1. Imagine that each player i can "decide" where to point in the construction of graph G_t at each round t, as a function of where everyone else is pointing.

Proof (Sketch)

1. Imagine that each player i can "decide" where to point in the construction of graph G_t at each round t, as a function of where everyone else is pointing.
2. We'll conclude that it is always in player i's best interest to point to his favorite good among the ones remaining.

Proof (Sketch)

1. Imagine that each player i can "decide" where to point in the construction of graph G_t at each round t, as a function of where everyone else is pointing.
2. We'll conclude that it is always in player i's best interest to point to his favorite good among the ones remaining.
3. Why might player i *not* want to point to his most preferred good?

Proof (Sketch)

1. Imagine that each player i can "decide" where to point in the construction of graph G_t at each round t, as a function of where everyone else is pointing.
2. We'll conclude that it is always in player i's best interest to point to his favorite good among the ones remaining.
3. Why might player i *not* want to point to his most preferred good?
4. Fear: If he points to a less preferred good, he gets it; if he points to his most preferred good, he doesn't, and his previous opportunity disappears.

Proof (Sketch)

1. Imagine that each player i can "decide" where to point in the construction of graph G_t at each round t, as a function of where everyone else is pointing.
2. We'll conclude that it is always in player i's best interest to point to his favorite good among the ones remaining.
3. Why might player i *not* want to point to his most preferred good?
4. Fear: If he points to a less preferred good, he gets it; if he points to his most preferred good, he doesn't, and his previous opportunity disappears.
5. But that can't happen...

Proof (Sketch)

1. Fixing the edges of the other players, consider the set of goods that agent i can get today if he points to them.

Proof (Sketch)

1. Fixing the edges of the other players, consider the set of goods that agent i can get today if he points to them.
2. These are the goods that form paths leading to agent i (will form cycles if agent i points to them). Call them "agent i's choice set"

Proof (Sketch)

1. Fixing the edges of the other players, consider the set of goods that agent i can get today if he points to them.
2. These are the goods that form paths leading to agent i (will form cycles if agent i points to them). Call them "agent i's choice set"
3. Agent i's choice set can only increase!

Proof (Sketch)

1. Fixing the edges of the other players, consider the set of goods that agent i can get today if he points to them.
2. These are the goods that form paths leading to agent i (will form cycles if agent i points to them). Call them "agent i's choice set"
3. Agent i's choice set can only increase!
4. As goods are removed, other agents might now point to agent i (choice set increases)...

Proof (Sketch)

1. Fixing the edges of the other players, consider the set of goods that agent i can get today if he points to them.
2. These are the goods that form paths leading to agent i (will form cycles if agent i points to them). Call them "agent i's choice set"
3. Agent i's choice set can only increase!
4. As goods are removed, other agents might now point to agent i (choice set increases)...
5. And nothing is removed, since all such goods are part of paths leading to agent i, so are not part of cycles not involving agent i.

Proof (Sketch)

1. Fixing the edges of the other players, consider the set of goods that agent i can get today if he points to them.
2. These are the goods that form paths leading to agent i (will form cycles if agent i points to them). Call them "agent i's choice set"
3. Agent i's choice set can only increase!
4. As goods are removed, other agents might now point to agent i (choice set increases)...
5. And nothing is removed, since all such goods are part of paths leading to agent i, so are not part of cycles not involving agent i.
6. Tada!

Thanks!

See you next class — stay healthy!