6.896: Topics in Algorithmic Game Theory

Lecture 18

Constantinos Daskalakis
Overview

- Social Choice Theory
- Gibbard-Satterwaite Theorem
- Mechanisms with Money (Intro)
- Vickrey’s Second Price Auction
- Mechanisms with Money (formal)
Social-Choice Preliminaries
Social Choice Theory

Setting:

\( A \): Set of alternatives ("candidates")

\( I \): Set of \( n \) voters

\( L \): Preferences on \( A \); usually this is the set of total orders on \( A \)

*Social Welfare Function*: \( f : L^n \rightarrow L \)

*Social Choice Function*: \( f : L^n \rightarrow A \)
Arrow’s Impossibility Theorem

Theorem [Arrow ’51]

*Every social welfare function on a set A of at least 3 alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.*

Proof: Last Lecture
Electing a President

- use a social choice function $f$

- ideally $f$ should satisfy the following properties:

  1. it should not be a *dictatorship*

    **Def:** A social choice function $f$ is a *dictatorship* if there exists some voter $i$ such that
    \[ f(<_1, <_2, \ldots, <_n) = \text{top}(<_i); \]
    Such voter $i$ is called the *dictator* of $f$.

  2. it should not be susceptible to *strategic manipulation*

    **Def:** $f$ can be strategically manipulated by voter $i$ if there exist preferences $<_1, <_2, \ldots, <_n$ and $<_i'$ such that
    \[ f(<_1, \ldots, <_i', \ldots, <_n) = a <_i a' = f(<_1, \ldots, <_i', \ldots, <_n) \]
    i.e. $i$ can elect a preferable candidate by lying
    If $f$ cannot be manipulated it is called *incentive compatible*. 
**Monotonicity**

**Def:** $f$ is *monotone* iff

$$f(<_{1},...,<_{i},...,<_{n}) = a \neq a' = f(<_{1},...,<_{i}',...,<_{n}) \Rightarrow \begin{cases} a' <_{i} a \\ \text{and} \\ a <_{i'} a' \end{cases}$$

i.e. if the outcome changes from $a$ to $a'$ when $i$ changes his vote from $>_{i}$ to $>_{i'}$, then it must be because the swing voter $i$ also switched his preference from $a$ to $a'$

**Proposition:**

($f$ is incentive compatible) iff ($f$ is monotone)

**Proof:** Immediate by definition.
Gibbard-Satterthwaite Thm
Gibbard-Satterthwaite Theorem

Theorem:

If \( f \) is an incentive compatible social choice function \textit{onto} a set of alternatives \( A \), where \(|A| \geq 3\), then \( f \) is a dictatorship.

Remark: “onto” is important; if \(|A|=2\) then the majority function is both incentive compatible and non-dictatorship.

Proof Idea: Suppose \( f \) is both incentive compatible and non-dictatorship. Use \( f \) to obtain a social welfare function \( F \) that satisfies unanimity, independence of irrelevant alternatives and non-dictatorship, which is impossible by Arrow’s theorem.
Proof of the GS theorem

*From the social choice function* $f$ *to a social welfare function* $F$

Notation: If $S \subseteq A$, and $\prec \in L$, we denote by $\prec^S$ the preference obtained from $\prec$ by moving all elements of $S$ to the top of $\prec$.

*Example:* $S = \{a, b\}$, and $x < a < y < b < z$ then $x \prec^S y \prec^S z \prec^S a \prec^S b$.

Definition of $F(\prec_1, \prec_2, \ldots, \prec_n) := \prec$

$a < b \iff f(\prec_1 \{a, b\}, \prec_2 \{a, b\}, \ldots, \prec_n \{a, b\}) = b$

**Claim 1:** $F$ is a social welfare function.

**Claim 2:** $F$ satisfies unanimity, IIA, and non-dictatorship.
Proof of the GS theorem (cont.)

**Lemma:** For any $S$, $<_1, <_2, \ldots, <_n$, $f( <_1^S, <_2^S, \ldots, <_n^S) \in S$.

**Proof:** hybrid argument, on board.

**Claim 1:** $F$ is a social welfare function.

**Proof:** By direct application of lemma, $F$ is a total order and it is anti-symmetric.

Transitivity?

Suppose that $a < b < c < a$ (*).

W.l.o.g. suppose that $f( <_1^{\{a, b, c\}}, <_2^{\{a, b, c\}}, \ldots, <_n^{\{a, b, c\}}) = a$.

Hybrid argument: by sequentially changing $<^{\{a, b, c\}}$ to $<^{\{a, b\}}$ argue that $f( <_1^{\{a, b\}}, <_2^{\{a, b\}}, \ldots, <_n^{\{a, b\}}) = a$, contradiction to (*).
Proof of the GS theorem (cont.)

Claim 2: F satisfies unanimity, IIA, and non-dictatorship.

Proof:

unanimity, IIA on board

non-dictatorship: 2 points
Mechanisms with Money
Going beyond the GS obstacle

- The GS theorem applies to the setting where voters declare ordinal preferences over the alternatives, rather than cardinal preferences.

- What if the voters assign a “score” to each alternative?

\[ v_i : A \to \mathbb{R} \]

\[ v_i(a) : \text{value of alternative } a \text{ for voter } i, \text{ in terms of some currency} \]

- Voter’s utility if alternative \( a \) is chosen and money \( m_i \) is given to him

\[ u_i = v_i(a) + m_i \]

\text{quasi-linear preferences}
Example 1: Auctioning off a single item

- each bidder $i$ has value $w_i$ for the item

- alternatives $A = \{1 \text{ wins}, 2 \text{ wins}, \ldots, n \text{ wins}\}$

- for all $i$:
  \[
  v_i(\text{$i$ wins}) = w_i \\
  v_i(\text{$j \neq i$ wins}) = 0
  \]

- suppose we want to implement the social choice function that gives the item to the bidder with the highest value for the item

- unfortunately we don’t know the $w_i$’s

- want to cleverly design the payment scheme to make sure that the social choice cannot be strategically manipulated
Example 1: Auctioning off a single item (cont)

- first attempt: no payment

- second attempt: pay your bid

- third attempt: *Vickrey’s second price auction*

  the winner is the bidder $i$ with the highest declared value $w_i = \max_j w_j$

  non-winners pay 0, and the winner pays $\max_{j \neq i} w_j$

**Theorem (Vickrey):** For all $w_1, w_2, \ldots, w_n$ and $w_i'$, let $u_i$ be bidder $i$’s utility if she bids her true value $w_i$ and let $u_i'$ be her utility if she bids an untrue value $w_i'$. Then $u_i \geq u_i'$. 
General Framework
Mechanisms with Money

Setting:

\( A : \) Set of alternatives ("candidates")

\( I : \) Set of \( n \) players

\[ v_i : A \rightarrow \mathbb{R} \]

valuation function of player \( i \)

\[ v_i \in V_i \subseteq \mathbb{R}^A \]

set of possible valuations

Def: A direct revelation mechanism is a collection of functions \((f, p_1, \ldots, p_n)\) where

\[ f : V_1 \times \ldots \times V_n \rightarrow A \]

is a social choice function

and

\[ p_i : V_1 \times \ldots \times V_n \rightarrow \mathbb{R} \]

is the payment function of player \( i \).
Def: A mechanism \((f, p_1, \ldots, p_n)\) is called \textit{incentive compatible}, or \textit{truthful}, or \textit{strategy-proof} iff for all \(i\), for all \(v_1 \in V_1, \ldots, v_n \in V_n\) and for all \(v'_i \in V_i\)

\[
v_i(a) - p_i(v_i, v_{-i}) \geq v_i(a') - p_i(v'_i, v_{-i})
\]

\(a = f(v_i, v_{-i})\)

\(a' = f(v'_i, v_{-i})\)

utility of \(i\) if he says the truth

utility of \(i\) if he lies

\(i.e.\) no incentive to lie!

but isn’t it too good to be true?