Algorithmic Game Theory, Summer 2015 Lecture 8 (4 pages)

Mechanism Design Basics

Instructor: Xiaohui Bei

Last lecture talks about the theory of *social choice*, which is about deciding a social outcome given each individual's preferences over outcomes. From this lecture on, we move into the theory of *mechanism design*, which studies how to *implement* the desired social choices in a strategic setting. That is, we assume that the preference of each individual is private information, and that they will act strategically in a game theoretic sense. This is the real "meat" of the second half of this course.

1 Model

An environment E of a mechanism design problem consists of

- a finite set of possible **outcomes** or **alternatives** O
- a set of agents $N = \{1, 2, \ldots, n\}$
- a set of utility functions $\mathbf{U} = U_1 \times \cdots \times U_n$ of agents, in which each $u_i \in U_i$ is of form $u_i : O \mapsto \mathbb{R}$.

Suppose that we are also given a choice rule $f : U \mapsto 2^O$ that maps each possible combination of utility functions to a set of acceptable outcomes.

Definition 8.1 (Mechanism)**.** *A* **mechanism** *for a given environment $E = (O, N, \mathbf{U})$ is a pair $(\mathbf{X}, g)$, where*

- *$\mathbf{X} = X_1 \times \cdots X_n$, where X_i is the set of strategies available to agent i, and*
- *$g : X_1 \times \ldots \times X_n \mapsto O$ is an outcome function that maps each possible combination of strategies to an outcome.*

The idea is that the designer gets to specify a game with sets of strategies $X_1, \ldots, X_n$ for each agent and an outcome function g, such that $u_i^G(x_1, \ldots, x_n) = u_i(g(x_1, \ldots, x_n))$. That is, the utility that agent i receives in this game for a combination of strategies is equal to his utility from the outcome associated to this combination.

2 Revelation Principle

The goal of mechanism design is to pick a mechanism that can guide rational agents with private information to behave in a desired way. Another way to look at this setting is that, one wants to design a game that *implements* a particular social choice function in equilibrium, given that the designer does not have information about agents' preferences.

So what does it mean to implement a social choice function? There are many different type of implementations. Among them, the strongest one is called *implementation in dominant strategies*.

Definition 8.2 (Implementation in dominant strategies)**.** *Given an environment $E = (O, N, \mathbf{U})$, a mechanism $(\mathbf{X}, g)$ is an* **implementation in dominant strategies** *of a social choice function $C : \mathbf{U} \mapsto O$, if for any possible utility functions $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbf{U}$, the game has a dominant strategy equilibrium, and in any such equilibrium a^*, we have $g(a^*) = C(\mathbf{u})$.*

There are certain advantages in designing an implementation in dominant strategies mechanism. First, it is easy for an agent to decide what to do in such a mechanism: just play the obvious dominant strategy. Second, the designer can predict the outcome of the mechanism by assuming that every agent plays their dominant strategies, which makes the mechanism much easier to analyze. Notice that although dominant strategy equilibria rarely occur in natural games, in a mechanism design problem, we have the flexibility to design our game to ensure that it has such nice property.

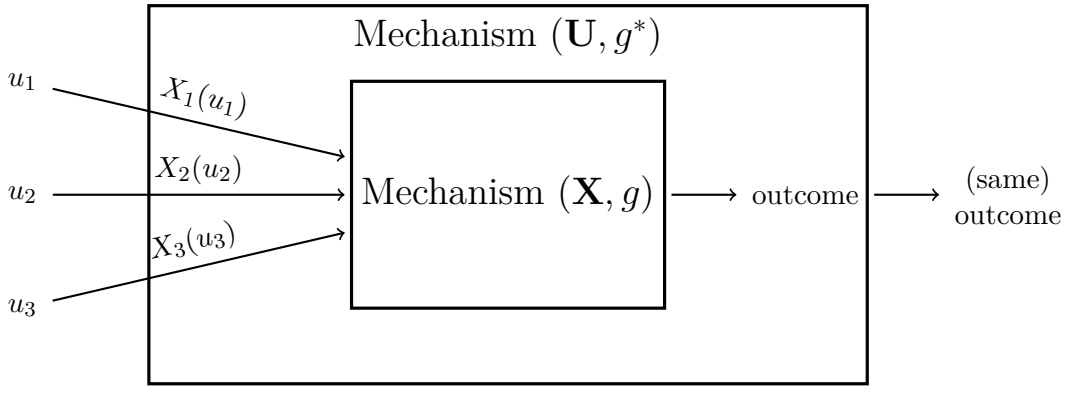
Furthermore, the next theorem says that implementation in dominant strategies mechanisms can enjoy an addition property, namely that we can convert the mechanism into a *truthful* mechanism that implements the same social choice function.

Notation. Let $\mathbf{u} = (u_1, \ldots, u_n)$ be a n-dimensional vector, we will use $\mathbf{u}_{-i}$ to denote the $n-1$-dimensional vector by removing ith coordinate from vector $\mathbf{u}$, i.e., $\mathbf{u}_{-i} = (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)$, and the original vector $\mathbf{u}$ can now be written as $\mathbf{u} = (u_1, \ldots, u_n) = (u_i, \mathbf{u}_{-i})$. Similarly, for $\mathbf{U} = U_1 \times \ldots \times U_n$, we will denote $\mathbf{U}_{-i} = U_1 \times \cdots U_{i-1} \times U_{i+1} \times \cdots \times U_n$.

Definition 8.3 (Incentive Compatible (IC))**.** *A mechanism (X, g) for a given environment $E = (O, N, \mathbf{U})$ is* **incentive compatible***, or* **truthful***, if the strategy space $\mathbf{X} = \mathbf{U}$, and for which $(u_1, \ldots, u_n)$ is a dominant strategy equilibrium in the game with utilities $(u_1, \ldots, u_n)$.*

Theorem 8.4 (Revelation Principle)**.** *If a mechanism $(\mathbf{X}, g)$ implements a social choice function C in dominant strategies, then there exists an incentive compatible mechanism that implements C.*

Proof. The proof uses a simulation argument. Assume that for any utility functions $(u_1, \ldots, u_n)$, $(X_1(u_1), \ldots, X_n(u_n)) \in \mathbf{X}$ is a dominant strategy equilibrium of the original mechanism. Then define a new mechanism $(\mathbf{U}, g^*)$ where $g^*(u_1, \ldots, u_n) = g(X_1(u_1), \ldots, X_n(u_n))$. Since $(X_1(u_1), \ldots, X_n(u_n))$ is a dominant strategy equilibrium, for any $\mathbf{x}_{-i}, x_i'$ we have $u_i(x_i', \mathbf{x}_{-i}) \geq u_i(x_i, \mathbf{x}_{-i})$. In particular this is true for $\mathbf{x}_{-i} = X(u_{-i}')$ and $x_i' = X(u_i')$ with any $\mathbf{u}_{-i}' \in \mathbf{U}, u_i' \in U_i$. This gives the definition of incentive compatibility of the new mechanism. □



Clearly when designing a mechanism that implements a certain social choice function, we want that participants have no difficulties in choosing what to act: rational players will want to report us truthfully their private preferences. And revelation principle tells us that in order to achieve this, it is enough to just implement the social choice function in dominant strategies.

Given the superiority of implementations in dominant strategies. Naturally the next question to ask is: does every social choice function admit such an implementation? Unfortunately, similar to what we have in social choice theory, the next theorem provides us a negative answer.

Theorem 8.5 (Gibbard-Satterthwaite)**.** *Let $C : \mathbf{U} \mapsto O$ be a social choice function with $|O| \geq 3$ and that C is onto, if $\mathbf{U}$ is the set of all possible utility function combinations, then C can be implemented in dominant strategies if and only if C is dictatorial.*

We provide an informal sketch of the proof. The idea is to reduce the problem to Arrow's impossibility theorem for social choice functions. Assume that a social choice function C can be implemented in dominant strategies. Then it can also be implemented by an incentive compatible mechanism. This means

$$u_i(C(u_i, u_{-i})) \geq u_i(C(u_i', u_{-i}))$$

holds for any $u_i, u_i' \in U_i$ and $u_{-i} \in \mathbf{U}_{-i}$. Assume for simplicity that all utility relations are strict. It can be shown that C must be weakly Pareto efficient and monotonic, otherwise there will exist some scenarios where some agent can obtain a higher utility by misreporting his preference. Then by Arrow's impossibility theorem, C must be either binary or dictatorial.

3 Quasilinear Environments

The negative result of Theorem 8.5 indicates that we cannot hope to design dominant strategies implementation (or incentive compatible) mechanisms for general utility functions. Hence people turn to look at special subclasses of utility functions for which a dominant strategies implementation always exists. The one that we will be studying in this lecture is the *quasilinear environment*. Compare to the general outcomes and utility functions, a quasilinear environment has the following special properties:

- the set of outcomes has form $O = A \times \mathbb{R}^n$, where A is the set of possible allocations, and $\mathbb{R}^n$ represents the payments that agents have to pay (or receive) for a given allocation,
- all possible utility functions are of form $u_i(a, (p_1, \ldots, p_n)) = v_i(a) - p_i$ for any $(a, (p_1, \ldots, p_n)) \in O$, where $v_i(a)$ is the value of allocation a to agent i, and p_i is the amount of money that agent i needs to pay (or receive if p_i is negative).
- the social choice function $C(\mathbf{u}) = (a, (p_1, \ldots, p_n))$ maps utility functions $\mathbf{u}$ to the allocation a that maximize $\sum_i v_i(a)$. Such social choice functions are called *efficient*.

3.1 Example: Single-Item Auctions

The simplest example in quasilinear environment is probably the *single-item auctions*: assume that a seller wants to auction a single item and there are n potential bidders. Each bidder i has a private value w_i, which represents the maximum amount of money that this bidder is willing to pay for this item. We focus on a special class of auction formats called *sealed-bid auctions*. Such auctions follow the following procedure:

- Each bidder i submits a bid b_i in a sealed envelope to the auctioneer, hence his communication with the auctioneer is private.
- The auctioneer decides the winner of this item based on these bids.
- The auctioneer decides the amount of money that the winner needs to pay for this item.

It is easy to model such setting as a mechanism design problem in a quasilinear environment: the set of allocations here is the set of possible winners, i.e., $A = 1, \ldots, n$, and for each bidder i, his valuation function is $v_i(a) = w_i$ for $a = i$ and $v_i(a) = 0$ for all $a \neq i$. The goal is to design a

mechanism that implements an efficient social choice function truthfully. Obviously in order to maximize $\sum_i v_i(a)$, we should always let the winner be the bidder that has the highest valuation w_i. The problem is how to design an appropriate payment rule, such that all rational bidders want to submit their private values w_i truthfully as their bids. This is not a trivial task. For example, if we do not charge any money and simply give this item for free to the bidder with the highest bid, then every bidder, regardless he values the item, would want to win this auction. And the auction will quickly turn into a game of "who can name the highest number".

First-Price Auctions A natural choice is to ask the winner to pay its bid. Such auction is called the *first-price auction*, which is common in practice. However, it is not hard to see that this is not an incentive compatible mechanism. Consider the following simple example: there are two bidders with private valuations $w_1 = 1$ and $w_2 = 2$. Assume that bidder 1 submits its bid $b_1 = 1$ truthfully. Now for bidder 2, if he submits his bid truthfully, he will win the item but also pay a price of 2, hence getting a totally utility of 0. However, a better strategy for this bidder is to submit a lower value $w_1 < b_2 < w_2$, such that he will still win the item but with a lower price, resulting a net positive utility. Thus this auction is not incentive compatible.

Second-Price Auctions Now let's consider another type of auction: again let the winner be the bidder i with the highest bid b_i, but ask the winner to pay the price that equals to the second highest bid of all bidders $p = \max_{j \neq i} b_j$.

This is called the *second-price auction* (also known as the *Vickrey auction*), which is also used in many practice scenarios. Moreover, it turns out that this auction enjoys the important property of truthfulness.

Theorem 8.6. *A second-price auction is an incentive compatible mechanism.*

Proof. For any given bidder i with private valuation w_i, let $b^* = \max_{j \neq i} b_j$ be the highest bid among other bidders. Note that if bidder i wins the item, he will get a fixed utility of $w_i - b^*$. If he bids anything lower than b^*. We consider two scenarios:

(1) By bidding w_i, bidder i wins the item. In this case we have $w_i \geq b^*$, hence his utility will be $u_i = w_i - b^* \geq 0$. Now, if he bids anything that wins the auction, his utility remains unchanged at u_i. If he bids anything that loses the auction, his utility will be $u_i' = 0 \leq u_i$.

(2) By bidding w_i, bidder i loses. This implies $w_i \leq b^*$ and he he receives a utility of $u_i = 0$. If he bids anything that wins the auction, his utility would become $u_i' = w_i - b^* \leq 0 = u_i$. If he bids anything that loses the auction, his utility will remain at 0.

From the above discussions, we can see that in both cases, bidding the true valuation can give this bidder the highest utility that he can get. Hence truth-telling is a dominate strategy for this bidder. □

It is not a coincidence that we can find an incentive compatible mechanism for single-item auctions. In later lectures we will see that in a quasilinear environment, any efficient social choice function C can actually be implemented by an incentive compatible mechanism.

Recommended Literature

- Chapter 9.2.4, 9.3, 9.4 in the AGT book.
- Tim Roughgarden's lecture notes `http://theory.stanford.edu/~tim/f13/l/l2.pdf` and lecture video `https://youtu.be/z1QZqYuiGa8`