CS-621 Theory Gems October 11, 2012

Lecture 8

Lecturer: Aleksander Mądry *Scribes: Alina Dudeanu, Andrei Giurgiu*

1 Mechanism Design

So far, we were focusing on "static" analysis of games. That is, we considered scenarios in which the game and utilities of all the players are fixed and known and our task is only to predict possible outcomes of that game when some (or all) the players are acting rationally.

Today, we turn the tables: there is no predefined game, only players that have some utilities. However, the key point is that these utilities are *private*. That is, we have no access to them (we only know a universe they are coming from) – so, in particular, when players claim to have some utility function there is no way for us to know if they are telling the truth.

Our goal now is to design a game that compels players that are acting rationally (with respect to their private utilities) to choose an outcome that maximizes the *social welfare*, i.e., a one that maximizes the sum of (private) utilities of all the players. (Note that an outcome that maximizes the social welfare might not necessarily be optimal from the point of view of any particular player. So, the difficulty here is to ensure that the social-welfare outcome is still the preferable one for all the players and, furthermore, to do it in a way that does not even require us to know what their actual utilities are.)

2 Vickrey Auction

Let's start with a motivating example. Consider the following setup: we have one item to auction and n bidders, each of them has a private valuation v_i of the item. Our goal is to design a way of auctioning the item that ensures that the item goes to the person that values it most – this can be seen as corresponding to maximization of social welfare.

Note that in the auction setting that is probably most familiar to us, an auctioneer cares for something else: to maximize his/her own revenue. This is not the case here, and there actually are real-world situations where a social-welfare objective makes sense. For example, when the government is auctioning radio frequencies, its main goal (instead of just making profit) is to ensure that whoever gets these frequencies will be willing and capable of utilizing them to greatest extent. Another real-world scenario is auctioning blocks of unused IP addresses (that due to running out of IP space become a sought-after commodity). The not-for-profit body that oversees this process is obviously interested in giving them to an organization that needs it (and thus values it) most and getting a revenue is actually undesirable.

Before we proceed, let us cast the above scenario into a more formal framework. The way we will view an auction is as a process in which first, each bidder i submits a bid b_i. Next, there is a *public* (i.e., known to every bidder before placing his/her bids) *outcome function* $f(b_1, \ldots, b_n) = (\bar{i}, p_1, ..., p_n)$ that based on these bids determines the winner $\bar{i}$ of the auction (to whom the item is given), as well as, payments p_i that for each bidder i that he/she has to pay. (Note that, in principle, we allow here situations in which bidder has to pay some amount even if he/she did not win an item.)

Now the resulting utility u_i of bidder i is defined as

$$u_i(\bar{i}, p_1, ..., p_n) = \begin{cases} v_i - p_i & \text{if } i = \bar{i}, \\ -p_i & \text{otherwise.} \end{cases}$$

So, our task here is to choose the outcome function f in such a way that bidders that are rational with respect to their utility functions (and f) are compelled to favor the outcome that maximizes the social welfare, i.e., gives out the item to the bidder that has maximum valuation v_i of it. Also, to make sure that any rational bidder will be interested in participating in the auction, we impose an additional condition that every bidder can always make his expected utility non-negative. In our case, this boils down to requiring that only the winner can be charged with non-zero payment, i.e., p_i is zero whenever $i \neq \bar{i}$.

At this point, the question is: what is the right choice of the winner and what payment should he/she be charged with?

Given that we are not interested in getting a revenue and want to just give the item to the bidder that has the highest valuation for it, a tempting approach would be to always give the item to the highest bidder and not charge anyone. The hope here would be that the bids will reflect the actual valuations and thus this indeed will lead to the desired outcome.

Unfortunately, it is easy to see that such approach will fail miserably in this setting. As none of the bidders is required to be truthful about their private valuations, the rational strategy is for every bidder to just lie by bidding $+\infty$ irregardless of the actual valuation. Clearly, that is not the right solution.

One natural attempt to fixing the above over-bidding problem is to make the bidders accountable for their bids. That is, one could consider so-called *first-price auction* in which still the highest bidder gets the item (with ties broken arbitrarily), but the payment of the winner has to be equal to his/her stated bid. This prevents over-bidding, as winning an auction with inflated bid results in negative utility, but leads to an opposite problem - underbidding.

Namely, in this case nobody has an incentive to bid his/her true valuation of the item, as doing so guarantees zero utility (no matter if the item is won or not). So, the resulting dynamics would be that each bidder tries to underbid in hope that the resulting bid will still be high enough to win the auction (provided the original valuation is sufficiently large), while leaving some positive margin of utility in case of the win. As no bidder has any prior information about the valuations of the other bidders, this dynamics is completely unpredictable and impossible to analyze within our framework (i.e., without any assumptions on the priors of the bidders). Even more importantly, as an auctioneer we would never be sure if the resulting outcome of the auction is indeed optimal from social welfare point of view, or just some bidder was more lucky with his/her choice of the bid. This motivates looking for a better solution.

The crucial insight here comes from considering what would happen in first-price auction if the valuations of the bidders were actually public. It is not hard to see that in this case, if i_k is the bidder with the k-th largest valuation then the "best" underbid for each bidder i_k is the valuation $v_{i_{k+1}}$ of the next bidder in this ordering. In fact, one can show (we essentially prove it in Lemma 1) that such bids constitute a dominant strategy in this setting.

The above observation motivates using so-called *second-price (or Vickrey) auction*, in which the item is still awarded to the highest bidder, but the payment of this bidder is equal to the *second* highest bid. As it turns out, once we make this modification, the game-theoretic properties of the resulting auction improve dramatically. In particular, as we prove below, this auction is *incentive compatible (IC)*, i.e., bidding the true valuation is a dominant strategy for all the bidders.

Lemma 1 *Second-price auction is incentive compatible (IC).*

Proof Let us focus on the perspective of a bidder i in this auction and, for the sake of the argument, assume he/she submits a bid b_i that is different than his/her true valuation v_i.

Let us first consider the case of $b_i > v_i$. If someone outbids the bidder i then he/she could have as well bid $b_i = v_i$, since he/she would not win anyway. On the other hand, if he/she does win with this bid, then the only situation in which just bidding v_i would not lead to winning too would be if some other bidder, say j, had his/her bid below b_i, but still above v_i. However, in this case, bidder i will be forced to pay more than his/her valuation, and the resulting utility will be negative. So, we see that bidding v_i is always not worse (and sometime actually better) than bidding $b_i > v_i$.

Now, to consider the complementary case of $b_i < v_i$. If i wins with this bid, then so would he/she with bidding v_i and obviously the payment would be the same. On the other hand, if he/she loses, then let us consider the winning bid b_j of some bidder j. If $b_j \geq v_i$ then bidding v_i would not make any difference in resulting utility (it still would be zero). However, if b_j is bigger than b_i, but smaller than v_i then actually bidding v_i would lead to winning and thus getting positive utility.

So, we see that indeed bidding exactly v_i always leads to the best outcome, irrespectively of other bids, and sometime deviating from this bid can actually lower the resulting utility. This means that bidding one's own true valuation is a (strictly) dominant strategy, as desired.■

Note that one consequence of second-price auction being incentive compatible is that we, as an auctioneer, are certain that as long as all the bidders are rational they are bound to bid truthfully. As a result, we can be sure that the item is indeed allocated in socially optimal way, i.e., it always is given to the bidder that values it most. So, by setting up this auction we managed to achieve quite remarkable feat. We managed to leverage the bidders' own rationality to make them disclose to us their private valuations and choose a socially optimal outcome even though from perspective of everyone but the winner, this outcome is very suboptimal.

3 Mechanism Design Without Money

After finding the solution for the auction problem above, it is natural to wonder for what other type of problems a similar solution can be obtained. Also, as we already mentioned, having to use money payments is undesirable in some scenarios, therefore we would like to investigate a possibility of doing without them.

To this end, let us first define more precisely our goal here. Once again, we will be interested in designing a game (mechanism). Let A be the set of its possible outcomes. There will be n players with private utilities $u_1, \ldots, u_n$, where each of these utilities comes from the same universe and utility u_i provides corresponding player's valuation of every possible outcome from A.

Now, the dynamics of the game is that each player i submits his alleged utility function u'_i (that might or might not be true) and then there is a public (i.e., known to everyone beforehand) function f that maps all u'_is into an outcome a of the game, i.e., $f(u'_1, \ldots, u'_n) = a$.

Clearly, function f – that we will call the *social choice function* – is the core of the game description and our goal is to choose it in a way that makes the resulting game have a dominant strategy $\bar{u}^*$ such that

$$f(\bar{u}^*) = \arg\max_{a \in A} \sum_i u_i(a),$$

i.e., we want that if all the players are rational then they are compelled to follow this strategy $\bar{u}^*$ and this strategy will lead to choosing an outcome a^* that maximizes the social welfare $\sum_i u_i(a)$. (Note that the social welfare is measured in terms of the private utilities of the players, not the submitted ones.)

To relate the above framework to our auction example from above, note that there one can think that the set A corresponds to all possible choices of the winner, and the utility of each player i is equal to his/her valuation v_i of the item if he/she wins the item and is zero otherwise. Then, submitting a bid can be viewed as declaration of having utility function with corresponding valuation, and maximizing the social welfare function corresponds exactly to giving the item to a player that values its most. (Note, however, that we do not have payments here.)

Before proceeding further, we note that although above we just require that there exists some dominant strategy $\bar{u}^*$ that leads to maximization of the social welfare, we can actually without loss of generality require that this strategy consists of each player being truthful about his/her utility function.

Lemma 2 (Revelation principle) *If there exists a social choice function f that facilitates a social-welfare-maximizing dominant strategy $\bar{u}^*$, then there also exists a social choice function f' that is incentive compatible, i.e., in which this dominant strategy is just truthful submission of everyone's utility functions.*

Proof Let us fix some game with social choice function f whose social-welfare-maximizing dominant strategy $\bar{u}^*$ is not truthful, i.e., $\bar{u}^* \neq \bar{u}$. Observe that whenever player i is rational, there is a fixed reasoning that leads him/her to play u_i^* given his/her private utility function is u_i, i.e., we can exactly model the way in which the players choose to lie. We can then consider a new game in which the social choice function f' first applies this reasoning to the submitted utility functions and then applies the function f to the output of that reasoning. It is not hard to see that in this new game being truthful constitutes a dominant strategy, as desired. ■

In the light of the above, from now on, we can always constraint ourselves to looking for mechanisms that are incentive compatible and this does not reduce the generality of our investigation.

Now, as we already argued, being able to come up with an incentive compatible social choice function for a large class of useful problems would be very powerful. Unfortunately, as the following theorem states, this goal is too idealistic and essentially in any interesting setting, such a function has to be necessarily not too useful.

Theorem 3 (Gibbard-Satterthwaite) *If the social choice function f is IC, is onto A and $|A| \geq 3$, then f is a dictatorship, i.e. $f(u'_1, \ldots, u'_n) = \bar{f}(u'_{\bar{i}})$, for some fixed $\bar{i}$.*

We did not prove this theorem in the class. The proof can be found, e.g, in [1]. Roughly speaking, the proof of it is based on application of Arrow's Impossibility Theorem, which (roughly) states that any "reasonable" voting system with at least three alternatives has to be a dictatorship. Gibbard and Satterthwaite showed that existence of a non-dictatorship social function as in the statement of their theorem, would also imply existence of non-dictatorship "reasonable" voting system. Thus, the desired impossibility statement follows.

4 Mechanism Design with Money

The Gibbard-Satterthwaite Theorem effectively kills our dreams of designing mechanism without money. So, we now turn our attention to the setting when payments are allowed.

In this setting, we again have n players and a set of possible outcomes A. Also, each player i has a private *preference* v_i (coming from some fixed universe) that ranks/evaluates from the perspective of that player all the possible outcomes in A.

Similarly to the previous case, the dynamics of the game is that each player i will declare first his/her alleged preference v'_i (that again might or might not be true) and there is a public function f (which we will now simply call a *mechanism*) that given the vector $(v'_1, \ldots, v'_n)$ of declared preferences produces an outcome a, as well as, a vector $(p_1, \ldots, p_n)$ of payments for all the players. Now, given the outcome and the payments, the utility u_i of player i is equal to his/her preference $v_i(a)$ of the obtained outcome minus the payment p_i that he/she needs to pay, i.e., $u_i(a) = v_i(a) - p_i$.

Our task, again, is to come up with the mechanism f that is incentive compatible and whose corresponding truthful strategy results in maximizing the social welfare, i.e., declaring $v'_i = v_i$ by all the players is a dominant strategy and in this case our mechanism should produce an outcome a^* such that

$$a^* = \arg\max_{a \in A} \sum_i v_i(a).$$

Note that the social welfare is based only on the (private) preferences v_i of the players and not on their utilities u_i (that include the effect of payments). This distinction is crucial, as in this way we are able to incentivize the players to the desired behavior by influencing their utility via money, while not affecting our objective function (social welfare).

4.1 The Vickrey-Clarke-Grove (VCG) Mechanism

After the unpleasant failure of mechanism design without money, one might also become skeptical about the mechanism design with money. Maybe the second-price auction is just one of very few tasks for which mechanism design is possible?

Fortunately, it turns out that, once money payments are allowed, there is a very elegant, versatile and essentially automatic way of obtaining good mechanisms: the Vickrey-Clarke-Grove (VCG) Mechanism.

To describe the VCG mechanism, let us fix some set of outcomes A, a set of n players, and their private preferences $v_1, \ldots, v_n$. We should first note that as we want our mechanism to be incentive compatible, the choice of outcome a' for a given vector of declared preferences $(v'_1, \ldots, v'_n)$ is already predefined. Namely, it has to be that

$$a' = \arg\max_{a \in A} \sum_i v'_i(a), \tag{1}$$

with arbitrary tie breaking. Otherwise, our mechanism would not be maximizing social welfare when players are truthful (which we want to be the case).

So, the only (but crucial!) design choice that we need to make is how to set the vector of payments $(p_1, \ldots, p_n)$. In the VCG mechanism, the payment p_i of player i, given the declared preferences $(v'_1, \ldots, v'_n)$, is

$$p_i(v'_1, \ldots, v'_n) := -\sum_{j \neq i} v'_j(a') + h_i(v'_{-i}), \tag{2}$$

where a' is the outcome given by (1) and $h_i(v'_{-i})$ is certain quantity called *Clarke's potential* that depends on the declared preferences v'_j of all the players *but* i (we will make it precise later).

To gain some intuition regarding this choice of payments, note that the utility of player i with respect to this payment becomes

$$u_i(a', v'_1, \ldots, v'_n) = v_i(a') + \sum_{j \neq i} v'_j(a') - h_i(v'_{-i}), \tag{3}$$

where, again, a' is given by (1).

Now, the crucial thing to notice is that, once we ignore the term $h_i(v'_{-i})$ (that does not depend on the choice of v'_i nor a' and thus is beyond control of player i), submitting $v'_i = v_i$ by player i makes the outcome a' chosen via (1) become exactly the outcome that maximizes this player's utility (3) (once all the other v'_j are fixed)! So, player i has no incentive to submit any other choice of v'_i than v_i - if he/she is truthful he/she is guaranteed to get maximum utility that is possible in this situation anyway. We see now that the key property of the choice of payments (2) is that it made our goal (getting an outcome that maximizes social welfare) and goal of every player (maximizing his/her utility) perfectly aligned. In the light of this, we can conclude with the following lemma.

Lemma 4 *For any choice of Clarke's potentials $h_1, \ldots h_n$, the VCG payment rule* (2) *results in an incentive compatible mechanism that maximizes the social welfare.*

4.2 Clarke's Pivot Rule

Although the VCG mechanism, as presented above, meets all the requirements of our model, it still has one shortcoming – it might make all the payments negative, i.e., make all the players receive money for their participation. Needless to say, having a mechanism that loses money is not ideal, so let us try to fix that. Ideally, we would like our mechanism to have two additional properties:

- **(Individual Rationality)** Rational players should always get a non-negative utility. In our setting, this means that submitting $v'_i = v_i$ should never result in a payment that is greater than the preference of the obtained outcome. (This ensures that rational players have incentive to participate in the mechanism.);
- **(No Positive Transfers)** No players is ever paid any money, i.e., all p_i are always non-negative;

To achieve these properties, we will use the crank that was not utilized so far in the VCG mechanism: Clarke's potentials $h_1, \ldots, h_n$. Specifically, we will set these potentials according to so-called *Clarke's Pivot Rule*:

$$h_i(v'_{-i}) := \max_{a \in A} \sum_{j \neq i} v'_j(a), \tag{4}$$

for each player i.

Note that this definition indeed depends only on the submitted preferences of all the other players except i. Furthermore, we can prove the following lemma.

Lemma 5 *The VCG mechanism with Clarke's pivot rule* (4) *is individually rational and, as long as, all v_is are non-negative, there is no positive transfers.*

Observe that we can guarantee no positive transfers only if all preferences are non-negative (i.e., all the players view the outcomes of the mechanism as potentially profitable to them). In fact, one can prove that this restriction is unavoidable.

Proof Individual rationality follows since

$$\max_a \sum_{j \neq i} v'_j(a) - \sum_{j \neq i} v'_j(a') \geq 0,$$

for any outcome a'.

To see that the no positive transfer property holds too when v_is are non-negative, note that

$$u_i(a', v'_1, \ldots, v'_n) = v_i(a') + \sum_{j \neq i} v'_i(a') - \max_{a \in A} \sum_{j \neq i} v'_j(a) \geq 0,$$

since a' is chosen so as to maximize $\sum_i v_i(a')'$, $v'_i = v_i$ when the player i is truthful, and - due to non-negativity of v_is - the maximum social welfare can only decreases when there is one less player in the game. ■

Finally, note that after application of Clarke's pivot rule, we can express the payment p_i of player i corresponding to an outcome a' as

$$p_i(v'_1, \ldots, v'_n) := -\sum_{j \neq i} v'_j(a') + \max_{a \in A} \sum_{j \neq i} v'_j(a).$$

This quantity has a very intuitive interpretation. Namely, note that when all the players are truthful, $\sum_{j \neq i} v'_j(a')$ becomes equal to the social welfare that all the other players get out of the game, while $\max_{a \in A} \sum_{j \neq i} v'_j(a)$ is the social welfare that these players would get if player i was not participating. So, the payment of player i is equal to the total loss in the social welfare of the other players that resulted from his/her participation.

4.3 Examples

Let us now take a look at two examples of application of the VCG mechanism.

4.3.1 The Vickrey Auction

We first show how the Vickrey/second-price auction that we introduced at the beginning of the lecture, can be obtained directly from the VCG mechanism. To this end, let us choose the set A of outcomes to be $A = \{1, \ldots, n\}$, with an outcome $a' = \bar{i}$ being just the identifier of the winner of the auction. The preferences of the users are functions of the form

$$v_i(\bar{i}) = \begin{cases} w_i, & \text{if } i = \bar{i} \\ 0, & \text{otherwise,} \end{cases}$$

where w_i is the private valuation of the item by player i.

The outcome and payment of the resulting VCG mechanism are given by

$$f(v'_1, \ldots, v'_n) = \left(\bar{i} = \arg\max_i v'_i,\ \ p_1, \ldots, p_n\right).$$

and

$$p_i(\bar{i}, v'_1, \ldots, v'_n) = -\sum_{j \neq i} v'_j(\bar{i}) + \max_{\bar{j}} \sum_{j \neq i} v'_j(\bar{j}),$$

for each player i.

Clearly, the winner is always the player that declares highest valuation/bid. Now, to understand the payments, note that when i is not the winner, the $\bar{j}$ that maximizes the second sum is exactly $\bar{i}$, so $p_i = 0$, as desired. (In other words, participation of player i in the auction did not influence the outcome and thus he/she does not owe anything.) Next, let us consider the case when i is the winner. The first sum will be equal to 0, because none of the other players wins. In the second sum, $\bar{j}$ will be the player with the second highest bid (as he/she would win if player i would not participate), so his valuation $v_{\bar{j}}$ is exactly what player i owes. Thus indeed we recovered the second-price auction.

4.3.2 Public Project

Now, consider a situation in which government wants to decide whether to build a public project that could benefit n different parties (i.e., each party i has a benefit w_i from having the project built). As the project is quite costly – et us say its cost is C – the government wants to go ahead with it only if the total benefit to all the parties is at least that large, i.e., only if $C \leq \sum_i w_i$. How can it be done, when the benefits w_i are private? (In particular, the parties might try to lie about their benefits just to encourage the government to go ahead with the project.)

To cast this problem into the VCG framework, let us set A to be $A := \{\text{Build}, \text{Not build}\}$. For every player i, let us define his/her preference to be

$$v_i(a) = \begin{cases} w_i & \text{if the project is built} \\ 0 & \text{otherwise.} \end{cases}$$

To make sure that we build only if the total benefit is bigger than the cost of the project, we introduce an additional dummy player, who has a negative benefit $-C$ if something is built and 0 otherwise.

It is not hard to see that the resulting mechanism will provide a solution for our task. However, interestingly, one can check that the only time a player owes something is when that player makes the difference between the project being built and not being built. This, in turn, means that unless $\sum_i w_i$ is exactly C, the total sum of payments from all the parties will not cover the cost C of building the project. (Again, one can show that this is in some sense unavoidable.)

References

[1] Noam Nisan, Tim Roughgarden Eva Tardos, and Vijay V. Vazirani, *Algorithmic Game Theory*, Cambridge University Press, 2007.