The Clarke Tax Algorithm

Michael A. Goodrich

October 25, 2005

1 Introduction

In these notes, we will introduce the Clarke Tax Algorithm which is a utility-based social choice mechanism. We will then give an example of the algorithm. We will then prove that this mechanism removes the incentive to lie about one's true preferences.

The description of the Clarke Tax Algorithm is taken from [2], but much of the discussion, the example, and the proofs were generated by me.

2 What Is It the Algorithm

The idea behind the Clarke Tax Algorithm is that a tax will be assessed to each individual. The amount of this tax depends on how the individual's utilities affect the rest of the group. Since people do not want to pay taxes, a big assessment will decrease an individual's overall utility. To get a better understanding of this statement, it is helpful to restrict the problem a bit. Formally, the Clarke Tax Algorithm (CTA) is applicable only when utilities are *quasilinear*. We will discuss the definition of a quasilinear utility shortly, but before we do so it is helpful to review the definitions of a social choice function and a mechanism.

Recall that a social choice function is a mapping from the set of possible individual agent types, Θ_i, to some societal outcome, X. Thus, the social welfare function is given by $f : \Theta_1 \times \Theta_2 \times \ldots \times \Theta_n \rightarrow X$, where n is the number of agents in society. In the case of the Clarke Tax algorithm, the set X consists of some publicly available good g that benefits the individuals in the society. For example, the agents may be getting together to decide whether to build a swimming pool and, if they decide to build it, how big of a pool to build. The social welfare function that we want is the one that maximizes the sum of each agent's utility over this public good,

$$g^* = \arg\max_g \sum_{k=1}^{n} v_k(g) \tag{1}$$

where $v_i(g)$ is how much agent i values the good g. Suppose that there is some benefit to an agent for lying about his or her preferences. For example, suppose that the agents in society have to pay some taxes to help pay for the good, and these taxes depend on what the agent reveals about his or her valuation true $v_i(g)$. In this case, the set X is a vector consisting of the selected good and the taxes that the agents pay, $X = [g, tax_1, \ldots, tax_n]$.

The Clarke Tax algorithm is a mechanism that tries to *implement* the social welfare function f in Equation (1) in truth dominant strategies. We will show that the Clarke Tax algorithm will select the same

g as the one identified by Equation (1). It does this by assessing these taxes in such a way that agents do not have an incentive to lie about their valuation.

To understand what this means, note that a strategy to agent i is to reveal something other than his or her true valuation. Let $v_i(g)$ denote agent i's true valuation, and let $\hat{v}_i(g)$ denote what the agent reveals about his valuation. Thus, the set of strategies available to agent i are $S_i = \{\hat{v}_i(g)\}$. Thus, we are trying to find a function $m : S_1 \times S_2 \times ... \times S_n \to X$ such that there are dominant strategies, $(s_1^*, s_2^*, \ldots, s_n*)$, of the "mechanism game" with

$$\forall i, \;\; s_i = v_i(g).$$

To avoid the Gibbard-Satterthwaite theorem, we will restrict our attention to preference patterns that follow a quasilinear utility function. A utility is quasilinear if the following conditions hold:

- The choices (or outcomes) are associated with a consequence that is made up of abstract consequences plus some "divisible numeraire (e.g., money)" [2, page 208]. Given our desired social choice mechanism, the social choice is over the set $X = \{(g, \pi_1, \pi_2, \ldots, \pi_n)\}$, where n is the number of agents in society, and where π_1 is the portion of the numeraire assigned to agent 1, π_2 the portion assigned to agent 2, and so on. The g is a consequence of the canonical form used in Arrow's theorem, but associated with this consequence is a division of, for example, money or taxes that must be paid by the individual agents when a particular choice is made by society.

- The utility of the choice $c = (g, \pi_1, \pi_2, \ldots, \pi_n)$ is the utility of the g plus the numeraire assignment. In other words, the utility of choice c for agent i is given by

$$u_i(c) = v_i(g) + \pi_i, \tag{2}$$

  where $v_i(g)$ is the valuation[1] of the canonical consequence g. This valuation is simply the utility of this outcome if the numeraire is ignored.

In what follows, $\hat{v}$ indicates the strategy that an agent reveals (i.e., a potentially fabricated valuation) and v indicates the true valuation. The Clarke tax algorithm is defined as follows:

> Every agent $i \in A$ simultaneously reveals his/her valuation $\hat{v}_i(g)$ for every possible g.
> The social choice is $g^* = \arg\max_g \sum_{i \in A} \hat{v}_i(g)$.
> Every agent is levied a tax: $tax_i(g^*) = \sum_{j \neq i} \hat{v}_j \Big(\arg\max_g (\sum_{k \neq i} \hat{v}_k(g)) \Big) - \sum_{j \neq i} \hat{v}_j(g^*)$.

Look at the definition of the tax carefully — it has some tricky notation. Recall that the notation $\arg\max(f(x))$ returns that value of x where $f(\cdot)$ achieves its maximum value. This means that the phrase

$$\hat{v}_j \Big(\arg\max_g (\sum_{k \neq i} \hat{v}_k(g)) \Big)$$

is read *the value of* $\hat{v}_j$ *at the argument chosen by the* $\arg\max$ *from*

$$\arg\max_g (\sum_{k \neq i} \hat{v}_k(g)).$$

[1]Note how the problem of interpersonal comparison of utilities is solved. Since everyone has a common currency, the numeraire, and since the numeraire π and the valuation v are summed to form the overall utility, the valuation v must be expressed in terms of the units of the numeraire. This means that everyone's utility units are expressed in the same currency and are therefore comparable.

This crazy equation evaluates the social welfare function (sum of individual utilities) for the entire group except for agent i at the option that would have been chosen if agent i's utilities had been ignored. In other words, if agent i had not been part of the group then the societal choice would be the one that maximized the social welfare function of this subgroup; evaluate the subgroup social welfare function at this choice. The tax is then defined as the difference between the subgroup social welfare function evaluated at the subroup's choice, $\arg\max_g(\sum_{k\neq i}\hat{v}_k(g))$, and the subgroup social welfare function evaluated at the entire group's choice, g^*. In other words, you pay taxes equal to the amount that the group lost because you participated. Quoting from [1, page878], "In the Clarke [Tax] mechanism agent i pays a tax equal to his effect on the other agents if he is pivotal to the project decision, and he pays nothing otherwise."

The idea is that if agent i's utilities cause the subgroup's welfare to decrease, then this agent should be assessed a tax proportional to this decrease. This assessment removes any incentive that agent i has to lie about his or her true valuation. Unlike the paying agents to reveal their true preferences as in the Vickrey auction, the Clarke tax algorithm introduces a tax that disincentives dishonesty. More specifically, in terms of our quasilinear utilities in Equation (2), $\pi_i = -tax_i(g)$.

3 An Example

Suppose that you are responsible for purchasing a storage area network for the university to be shared by the Computer Science, the Electrical and Computer Engineering, and the Mechanical Engineering departments. Suppose that we are considering three possible storage sizes, $g \in \{16TB, 32TB, 64TB\}$. Suppose that each agent's utility over the options is shown in Table 1. Notice that $g = 16TB$ will be

i	$v_i(g)$		
	16TB	32TB	64TB
CS	1	2	3
ECEn	3	1	2
ME	3.5	1	2
$\sum_i v_i(g)$	**7.5**	**4**	**7**

Table 1: The true values of each agent.

chosen if all agents reveal their true preferences.

Now, let's see what the tax to the CS department will be given these utilities. This is determined by looking at what would be chosen by the group if CS were not present. This is shown in Table 2. Since the input from CS did not change the fact that $g = 16TB$ is chosen, CS does not have to pay a tax.

Now, note that the CS department most prefers $64TB$ so we lie about our true preferences and reveal the values shown in Table 3. When this happens, $g' = 64TB$ is chosen by the group. However, the CS department has to pay a tax equal to (us Table 2 to compute this) $\sum_{i\neq \mathrm{CS}} v_i(g^*) - \sum_{i\neq \mathrm{CS}} v_i(g') = 2.5$ which exceeds the amount of benefit that CS acquired by lying, $v_{\mathrm{CS}}(g') - v_{\mathrm{CS}}(g^*) = 2$.

How much tax does the ECEn department pay? How about the ME department?

Is this algorithm coalition-proof?

i	$v_i(g)$		
	16TB	32TB	64TB
CS			
ECEn	3	1	2
ME	3.5	1	2
$\sum_{i \neq \text{CS}} v_i(g)$	**6.5**	**2**	**4**

Table 2: What would be chosen if CS did not have a say.

i	$v_i(g)$		
	16TB	32TB	64TB
CS	1	2	**4**
ECEn	3	1	2
ME	3.5	1	2
$\sum_i v_i(g)$	**7.5**	**4**	**8**

Table 3: The true values of each agent.

4 The Clarke Tax Algorithm is Truth Dominant

Since the utilities are quasilinear, $u_i(c) = v_i(g) - tax_i(g)$, we can (without loss of generality) denote $u_i(c)$ uniquely as $u_i(g)$. This means that we will use the abstract consequence to denote the choice. The utility of the real consequence is given by $u_i(g) = v_i(g) - tax_i(g)$ which says that the utility of choosing the abstract consequence is modulated by a real tax that reduces the overall utility of the choice.

Let g^* denote the choice that would be made if agent i reveals his or her true preferences, and let g' denote the choice that would be made if agent i lies about his or her true preferences. We will show that the utility for lying is no higher than the utility for telling the truth, and thereby conclude that there is not incentive to lie.

To this end, what is the difference between agent i's utility with and without the lie:

$$\begin{aligned}
u_i(g^*) - u_i(g') &= \Big(v_i(g^*) - tax_i(g^*)\Big) - \Big(v_i(g') - tax_i(g')\Big) && (3)\\
&= \Big(v_i(g^*) - v_i(g')\Big) - \Big(tax_i(g^*) - tax_i(g')\Big) && (4)\\
&= \Big(v_i(g^*) - v_i(g')\Big) - \Big(-\sum_{j\neq i} \hat{v}_j(g^*) + \sum_{j\neq i} \hat{v}_j(g')\Big) && (5)\\
&= \Big(v_i(g^*) + \sum_{j\neq i} \hat{v}_j(g^*)\Big) - \Big(v_i(g') + \sum_{j\neq i} \hat{v}_j(g')\Big) && (6)\\
&\geq 0. && (7)
\end{aligned}$$

The first two lines include the definition of utility and a regrouping. Equation 5 substitutes the definition of the tax from the previous section, and subtracts out the common term $\hat{v}_j\Big(\arg\max_g(\sum_{k\neq i} \hat{v}_k(g))\Big)$. The next line regroups, and then Equation 7 states that the social welfare function $\Big(v_i(g^*) + \sum_{j\neq i} \hat{v}_j(g^*)\Big)$

is bigger than the social welfare function $\left(v_i(g') + \sum_{j \neq i} \hat{v}_j(g')\right)$; were this not so, g' would have been chosen by the group instead of g^*.

Putting these equations together allows us to conclude that

$$u_i(g^*) - u_i(g') \geq 0$$

which says that any lying about preferences that causes a change from g^*, the choice when the truth is told, to g' cannot produce an increase in utility. Thus, there is no incentive to lie.

References

[1] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

[2] G. Weiss, editor. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, 1999.