Econ/Math C103

Matching Theory

Haluk Ergin

1 Introduction

We will study a number of basic matching models and algorithms from the seminal works of David Gale, Lloyd Shapley, and Herbert Scarf in the 1960's and 1970's. Although the early works were motivated by pure theoretical interest, more recently the theory has been applied to many real-life situations which eventually led to Al Roth and Lloyd Shapley receiving the Nobel prize on matching and market design in 2012. The real-life applications of these models range from entry level labor markets (matching of medical residents to hospitals, law interns to law firms, etc.), to public school choice matching (matching of high school students to public high schools in New York, Boston, etc.), and live-donor organ exchanges (kidney, liver, and lung exchanges). In these notes, we will focus on theory and not the applications. If you are interested in learning more about the applications, see the survey by Al Roth listed at the end of the lecture notes among optional readings.

2 The Marriage Market

The marriage market is a *two-sided* matching model in which there are two types of agents, the men and the women, referred to as the two-sides of the market. We will assume that each agent can only be matched to agents from the other side, or remain unmatched. Members of each side have strict preferences over being matched to different members of the other side or remaining unmatched.

Definition 1 A **marriage market** is a triple (M, W, R) consisting of disjoint finite sets of men $M = \{m_1, \ldots, m_n\}$, and women $W = \{w_1, \ldots, w_p\}$, and a preference profile $R = ((R_m)_{m \in M}, (R_w)_{w \in W}) = (R_{m_1}, \ldots, R_{m_n}, R_{w_1}, \ldots, R_{w_p})$, such that:

(a) $\forall m \in M : R_m$ is a linear order over $W \cup \{m\}$, and

(b) $\forall w \in W : R_w$ is a linear order over $M \cup \{w\}$.

Given a marriage market (M, W, R), for any agent $i \in M \cup W$, let R_{-i} denote the preference profile of all agents but agent i.

We next formulate the notion of a matching, where we impose our assumption that the marriage market is a *one-to-one* model, in the sense that everyone can be matched to at most one member from the other side.

Definition 2 Given a marriage market (M, W, R), a **matching** is a function $\mu : M \cup W \to M \cup W$ such that:

1. (a) $\forall m \in M$, $\mu(m) \in W \cup \{m\}$

   (b) $\forall w \in W$, $\mu(w) \in M \cup \{w\}$

2. $\forall i \in M \cup W$, $\mu(\mu(i)) = i$.

Under the matching μ, the match $\mu(m)$ of a man m is either a woman if $\mu(m) \in W$, or himself if $\mu(m) = m$, the latter interpreted as m being unmatched. Similarly under μ, each woman w can either be matched to a man if $\mu(w) \in M$, or be unmatched if $\mu(w) = w$. Condition 2 is just a consistency requirement saying that the match of the match of every agent i must be i.

Example 1 Consider a marriage market with two men $M = \{1, 2\}$, two women $W = \{a, b\}$, where R is given by the following preference profile:

R_1	R_2		R_a	R_b
a	a		2	2
b	b		1	b
1	1		a	1

Consider the matching μ, which matches 1 and b, and 2 and a. We can also describe μ by listing pairs that it matches, as in $(1b, 2a)$, or underlining the agents' matches in the preference profile:

R_1	R_2		R_a	R_b
a	$\underline{a}$		$\underline{2}$	2
$\underline{b}$	b		1	b
1	1		a	$\underline{1}$

Note that under μ, woman b is matched to man 1, but she would rather be unmatched than be matched to him. Assuming that individuals can unilaterally sever ties, we would not expect matching μ to be the outcome of the above marriage market because woman b would divorce man 1. We will say that a matching is individually rational if no one has an incentive to divorce their current partner.

Definition 3 Given a marriage market (M, W, R), a woman w is **acceptable** to a man m if $wR_m m$. Similarly, a man m is **acceptable** to a woman w if $mR_w w$. A matching μ is **individually rational (IR)** if

$$\forall i \in M \cup W : \ \mu(i) R_i i$$

i.e., everyone is matched to an acceptable partner or remains unmatched.

Example 2 Consider now the marriage market with two men $M = \{1, 2\}$ and two women $W = \{a, b\}$, where everybody is acceptable to everybody on the other side, and the preference profile R and the matching μ are given by:

R_1	R_2		R_a	R_b
a	b		1	$\underline{1}$
$\underline{b}$	$\underline{a}$		$\underline{2}$	2

Note that at μ, man 1 and woman a prefer each other to their partners. Therefore, they can get strictly better-off by breaking up from their current partners and rematching with each other. We will call a matching to be pairwise stable if no such pairs exist.

Definition 4 Given a marriage market (M, W, R) and a matching μ, a pair $(w, m) \in M \times W$ is a **blocking pair at** μ if $wP_m\mu(m)$ and $mP_w\mu(w)$. A matching μ is **pairwise stable** if there are no blocking pairs at μ.

Definition 5 Given a marriage market (M, W, R), a matching μ is **stable** if it is individually rational and pairwise stable.

We next define the core of a marriage market as the set of matchings μ such that no subgroup of agents $M' \cup W'$ can rematch within themselves and improve their positions.

Definition 6 A matching μ is in the **core** of the marriage market (M, W, R) if there are no $M' \subset M$, $W' \subset W$, and a matching μ' s.t.:

1. $\mu'(M' \cup W') = M' \cup W'$.
2. $\forall i \in M' \cup W'$: $\mu'(i) R_i \mu(i)$ and $\exists i \in M' \cup W'$: $\mu'(i) P_i \mu(i)$.

Condition 1 above can be read as: Under μ', the agents in the subgroup $M' \cup W'$ are matched within the subgroup. Condition 2 corresponds to saying that under the alternative matching μ', all of the agents in $M' \cup W'$ are weakly better-off, and at least one of them is strictly better-off.

A matching is Pareto efficient if there is no other matching that improves the position of all the agents.

Definition 7 Given a marriage market (M, W, R), a matching μ is **Pareto efficient** if there does not exist a matching μ' such that: (1) $\mu'(i) R_i \mu(i)$ for every $i \in M \cup W$, and (2) $\mu'(i) P_i \mu(i)$ for some $i \in M \cup W$.

A direct implication of the last two definitions is that every matching in the core of a marriage market is also Pareto efficient (just pick $M' = M$ and $W' = W$ in the definition of the core). It is also not difficult to see that if a matching μ is in the core then μ must be stable, because stability rules out only a subset of the deviations ruled out from the core. More interestingly as we show next, in a marriage market, any stable matching must be in the core.

Proposition 1 *Given a marriage market, a matching is in the core if and only if it is stable. In particular, all stable matchings are also Pareto efficient.*

Proof: Covered in class. □

A natural question is whether we can guarantee the existence of stable matchings, equivalently the nonemptiness of the core in light of the above Proposition. Gale and Shapley (1962) show that the answer is yes, by introducing an algorithm that produces a stable matching for any given marriage market.

The Men-Proposing Deferred Acceptance (DA) Algorithm

Given a marriage market (M, W, R), consider the following algorithm:

Step 1: Every man m proposes to his favorite acceptable woman. For each w, among the men who proposed to w, w places her favorite acceptable man tentatively in her waiting list, and rejects the others.

Step $k \geq 2$***:*** Men rejected at step $k-1$ propose to their next best acceptable women. For each w, the favorite acceptable man, among the new proposers and the one (if any) already in the waiting list of w from step $k-1$, is placed on her new waiting list and the rest are rejected.

The algorithm terminates when every man is either on a waiting list or has been rejected by every acceptable woman. Note that after being rejected, a man either proposes to the next acceptable woman one step down in his preference list, or if there are no such women (i.e., if he has been rejected by every acceptable women) he never proposes again. Therefore, the algorithm terminates in at most $|M| \times |W|$ steps. At the end, women get matched to the man in their final waiting list, the others (women with empty waiting lists and the men who are not on any waiting list) are left unmatched.

Note also that since the two sides of the market are entirely symmetric, the women-proposing version of the DA algorithm can be defined analogously.

Let's make a clarification in order to avoid a common misunderstanding. The DA algorithm is used here to describe a particular matching. We are not trying to suggest that the participants in the marriage market are indeed going through a dynamic process of proposals and rejections as described by the algorithm. You should perceive the algorithm as method to compute a matching, and no more.

Example 3 Consider the marriage market given by $M = \{1, 2, 3, 4, 5\}$, $W = \{a, b, c, d\}$, and

R_1	R_2	R_3	R_4	R_5
a	d	d	a	a
b	b	c	d	b
c	c	a	c	d
d	a	b	b	5

R_a	R_b	R_c	R_d
2	3	5	1
3	1	4	4
1	2	1	5
4	4	2	2
5	5	3	3

Everybody finds everybody in the other side of the market acceptable except 5 who does not find c acceptable.

Consider the men-proposing DA algorithm for the above marriage market:
Step 1: Men 1, 4, and 5 propose to woman a; a puts 1 in her waiting list and rejects the others. Men 2 and 3 propose to woman d; d puts 2 in her waiting list and rejects 3.
Step 2: Man 3 proposes to c who puts him in her waiting list. Man 4 proposes to d, who rejects 2 and puts 4 in her waiting list. Man 5 proposes to b, who puts him in her waiting list.
Step 3: Man 2 proposes to b, who rejects 5 and puts 2 in her waiting list.
Step 4: Man 5 proposes to d who rejects him.

The waiting lists at the end of each step during the algorithm are as follows:

	a	b	c	d
Step 1	1			2
Step 2	1	5	3	4
Step 3	1	2	3	4
Step 4	1	2	3	4

At the end of Step 4, men 1, 2, 3, 4 are all on a waiting list, and man 5 has been rejected by every acceptable woman. Therefore the algorithm terminates, producing the matching:

$$\mu_M = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ a & b & c & d & 5 \end{pmatrix}$$

Note that the waiting lists are only tentative up until the end of the DA algorithm, i.e., a man who is on a waiting list at some step can be rejected from that waiting list in subsequent steps of the algorithm, as is the case for 2 and 5 in this above example.

One can similarly compute the outcome of the women-proposing DA algorithm for the above marriage market:

$$\mu_W = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ d & a & b & c & 5 \end{pmatrix}$$

Note that all men weakly prefer μ_M to μ_W, and all women weakly prefer μ_W to μ_M. We will later show that this is the case in all marriage markets.

Given a marriage market (M, W, R), let μ_M and μ_W denote the outcomes of the men-proposing and women-proposing DA algorithms, respectively.

Theorem 1 *(Gale & Shapley (1962)) For any given marriage market, the outcomes of the DA algorithm μ_M and μ_W are stable. In particular, there exists a stable matching in every marriage market.*

Proof: Covered in class. □

Definition 8 A binary relation $\geq$ on a set X is a **partial order** if it is reflexive, transitive, and anti-symmetric. We write $x \leq y$ to denote $y \geq x$. We write $x > y$ to denote $x \geq y$ and not $x \leq y$ (similarly for $x < y$).

Define the binary relation $\geq_M$ on the set of matchings by

$$\mu \geq_M \nu \quad \Leftrightarrow \quad \forall m \in M : \mu(m) R_m \nu(m).$$

Note that $\geq_M$ is a partial order. Define $\geq_W$ similarly.

Proposition 2 *(Gale & Shapley (1962)) $\mu_M \geq_M \nu$ for any stable matching ν. Similarly, $\mu_W \geq_W \nu$ for any stable matching ν.*

Proof: Covered in class. □

Therefore, μ_M is called the **men-optimal** stable matching. Similarly, μ_W is called the **women-optimal** stable matching.

Note that the above proposition implies in particular that all the men weakly prefer μ_M to μ_W and all the women weakly prefer μ_W to μ_M. This observation can be generalized to show that the two sides of the market have *opposing interests* over all stable matchings: For *any* two stable matchings μ and ν, all the men weakly prefer μ to ν if and only if all the women weakly prefer ν to μ.

Proposition 3 *(Knuth (1976)) Given a marriage market (M, W, R), for any two stable matchings μ and ν:*

$$\mu \geq_M \nu \Leftrightarrow \nu \geq_W \mu$$

Proof: Covered in class. $\square$

Corollary 1 Given a marriage market (M, W, R), μ_W is the worst stable matching for men and μ_M is the worst stable matching for women.

2.1 Strategic Properties of the Optimal Stable Mechanisms

Throughout the discussion in this section, we will fix (M, W). The **men-optimal stable mechanism** is a function f^{MOSM} that associates the men-optimal stable matching $f^{MOSM}(R)$ to every preference profile R. Let $f_i^{MOSM}(R)$ denote i's match under the man-optimal stable matching $f^{MOSM}(R)$.

Theorem 2 *(Dubins & Freedman (1981), Roth (1982)) The men-optimal stable mechanism is strategyproof for men, i.e., for any preference profile R, for any man $m \in M$, and any preference R'_m:*

$$f_m^{MOSM}(R_m, R_{-m}) R_m f_m^{MOSM}(R'_m, R_{-m}).$$

Proof: Omitted. $\square$

We can similarly define the **women-optimal stable mechanism** f^{WOSM}. Then, the symmetric result holds, i.e. the women-optimal stable mechanism is strategyproof for women.

Example 4 This example shows that the men-optimal stable mechanism is not strategyproof for the women. Let $M = \{1, 2\}$ and $W = \{a, b\}$. Consider the following preference profile R where every agent finds those on the other side acceptable:

R_1	R_2
a	b
b	a

R_a	R_b
2	1
1	2

Note that the men-optimal stable matching at the above preference profile is $f^{MOSM}(R) = (1a, 2b)$. If woman a reports the preference ranking $R'_a : 2P'_a a P'_a 1$ instead of her true preference ranking R_a, then the men-optimal stable mechanism would return the matching $f^{MOSM}(R'_a, R_{-a}) = (1b, 2a)$ which is strictly better for a under her true preferences.

3 One-Sided Matching: The Roommates Problem

The roommates problem is a *one-sided* matching model, with one type of agent: students looking for roommates. Each student has preferences over being roommates with any other student and remaining unmatched.

Definition 9 A **roommates problem** is a pair (N, R) consisting of a finite set of agents $N = \{1, \ldots, n\}$, and a preference profile $R = (R_1, \ldots, R_n)$, such that R_i is a linear order over N for all $i \in N$.

We next formulate the notion of a matching, where we impose our assumption that the roommates problem is a *one-to-one* model, in the sense that every student can be matched to at most one other student as a roommate.

Definition 10 Given roommates problem (N, R), a **matching** is a function $\mu : N \to N$ such that $\mu(\mu(i)) = i$ for all $i \in N$.

If $\mu(i) \neq i$, we interpret $\mu(i)$ as i's roommate under the matching μ. As in before, $\mu(i) = i$ is interpreted as i being unmatched. The condition $\mu(\mu(i)) = i$ corresponds to the requirement that the roommate of the roommate of every agent i must be i. The notions of stability, core, and Pareto efficiency have natural counterparts in the roommates problem, that we state below for completeness.

Definition 11 Given a roommates problem (N, R), a matching μ is **stable** if it is:

1. **Individually Rational (IR)**: $\mu(i) R_i i$ for all $i \in N$.

2. **Pairwise Stable**: There are no two agents $i, j \in N$ such that $j P_i \mu(i)$ and $i P_j \mu(j)$.

Definition 12 A matching μ is in the **core** of a roommates problem (N, R) if there is no $N' \subset N$ and a matching μ' such that:

1. $\mu'(N') = N'$.

2. $\mu'(i) R_i \mu(i)$ for all $i \in N'$ and $\mu'(i) P_i \mu(i)$ for some $i \in N'$.

Definition 13 In a roommates problem (N, R), a matching μ is **Pareto efficient** if there does not exist a matching μ' such that $\mu'(i) R_i \mu(i)$ for all $i \in N$ and $\mu'(i) P_i \mu(i)$ for some $i \in N$.

The interpretations of the definitions are analogous to their counterparts in a marriage market. As in a marriage market, every matching in the core is Pareto efficient. The result about the equivalence of core and stability also carries over to the roommates problem.

Proposition 4 *Given a roommates problem, a matching is in the core if and only if it is stable. In particular, all stable matchings are also Pareto efficient.*

Proof: Exercise. □

The roommates problem is a model that is more general than the marriage market. To see this, note that any marriage market (M, W, R) can be "converted" to a roommates problem: Set $N = M \cup W$, and extend the preference R_i of every agent $i \in N$ to a preference $\tilde{R}_i$ over all of N in a way that he/she finds agents from the same side unacceptable. Then, it is easy to see that μ is a stable matching for the marriage market (M, W, R) if and only if it is a stable matching for the corresponding roommates problem $(N, \tilde{R})$ that we constructed.

The next example shows the non-emptiness of the core does not extend to the roommates problem.

Example 5 (Gale & Shapley (1962)) Consider a roommates problem (N, R) where $N = \{1, 2, 3\}$ and

R_1	R_2	R_3
2	3	1
3	1	2
1	2	3

The roommates problem has four matchings. The matching that leaves everyone unmatched is blocked by every pair of agents. In the other three matchings exactly one agent is unmatched. The matching that leaves 1 unmatched is blocked by 1 and 3; the matching that leaves 2 unmatched is blocked by 2 and 1; and the matching that leaves 3 unmatched is blocked by 3 and 2. Therefore there is no stable matching.

4 Assignment of Indivisible Objects: The Housing Market

The housing market is a model in which a finite set of agents, e.g. homeowners, each own an indivisible object, e.g. a house. Each agent can consume exactly one object, and has a strict preference ranking over all objects.

Definition 14 A **housing market** is a tuple (N, X, R, μ^E) where $N = \{1, \dots, n\}$ is a finite set of agents, X is a set of objects such that $|X| = n$, $R = (R_1, \dots, R_n)$ is a preference profile such that R_i is a linear order over X for all $i \in N$, and $\mu^E : N \to X$ is a bijection denoting the initial endowments.

The endowment function μ^E above summarizes which agent owns which object: Agent i owns object $\mu^E(i)$ for every $i \in N$.

Definition 15 Given a housing market (N, X, R, μ^E), an **assignment** is a bijection $\mu : N \to X$.

An assignment determines an allocation of the objects to the agents in a way that each agent receives exactly one object, and $\mu(i)$ denotes the indivisible object that agent i receives under assignment μ. Note that the initial endowment function μ^E is an assignment.

You may have noticed that the housing market and the marriage market models seem similar at least at an abstract level: In both models there are two disjoint sets and we are looking for methods to match members of the two sets to each other. However, there are two important differences that make the two models very different. First, in the marriage market model, both sides are agents and have preferences over the other side; in the housing market model, only one side consists of agents and the other side are merely objects to be consumed by the agents. Therefore, in the marriage market model we care about the welfare and strategic behavior of the two sides, whereas in the housing market, we only care about the welfare and strategic behavior of the agents' side. Second, the housing market model has an initial endowment assignment, which has no counterpart in the marriage market model. These differences will reflect in our formulations of Pareto efficiency, individual rationality, and the core. Note also that in the housing market model we are calling μ an assignment instead of a matching, to emphasize the distinction between the two models.

More minor differences between the two models are as follows. In the marriage market model, we haven't required $|M| = |W|$, we allowed for the possibility of staying unmatched, and also allowed agents to find some members of the other side unacceptable. In the housing market model, we are assuming that we have an equal number of objects and agents, we do not allow for not owning nor not consuming an object, and implicitly assume that all objects are acceptable to all agents. It is possible to relax these extra assumptions, but we will maintain them for expositional simplicity.

Next we introduce the notion of core in the housing market: an assignment is in the core if no subgroup of agents can reallocate the objects they own among each other and improve their positions.

Definition 16 An assignment μ is in the **core** of the housing market (N, X, R, μ^E) if there is no $N' \subset N$, and an assignment μ' such that:

1. $\mu'(N') = \mu^E(N')$.
2. $\forall i \in N'$: $\mu'(i) R_i \mu(i)$ and $\exists i \in N'$: $\mu'(i) P_i \mu(i)$.

The first condition above says that under the assignment μ', agents in N' are each assigned an object that is owned by an agent in N'. The second condition says that all agents in N' are getting weakly better-off and at least one agent in N' is getting strictly better-off under μ'.

We next introduce the notions of Pareto efficiency, and individual rationality for the housing market. An assignment is Pareto efficient if there is no other assignment that improves the position of all agents. An assignment is individually rational if all agents weakly prefer the object that they are assigned to the object that they initially own.

Definition 17 Given a housing market (N, X, R, μ^E), an assignment μ is **Pareto efficient** if there does not exist an assignment μ' such that: (1) $\mu'(i) R_i \mu(i)$ for every $i \in N$, and (2) $\mu'(i) P_i \mu(i)$ for some $i \in N$.

Definition 18 Given a housing market (N, X, R, μ^E), an assignment μ is **individually rational (IR)** if $\mu(i) R_i \mu^E(i)$ for every $i \in N$.

As in the sections before, it is easy to see that every assignment in the core is Pareto efficient and individually rational. A natural question is whether the core of a housing market is nonempty, and if so how can we find an assignment in the core. We will see that the answer to the first question is yes, and that the core of a housing market is actually always a *singleton*. We will show that the unique core allocation can be found by using the following algorithm introduced in Shapley and Scarf (1973). In that paper, Shapley and Scarf attribute this construction to David Gale. That is why the procedure is called Gale's Top-Trading Cycles (TTC) Algorithm.

Gale's Top-Trading Cycles (TTC) Algorithm

Given a housing market (N, X, R, μ^E), consider the following algorithm:

Step 1: Let every agent point to the owner of her most preferred object (possibly to herself). There is at least one cycle (including self-cycles). Assign every agent who is part of a cycle to her most preferred object. Let N^1 be the set of remaining agents and X^1 be the set of remaining objects. Note that $X^1 = \mu^E(N^1)$.

Step $k \geq 2$***:*** Let every agent in N^{k-1} point to the owner of her most preferred object in X^{k-1}. Assign every agent who is part of a cycle to her most preferred object in X^{k-1}. Let N^k be the set of remaining agents and X^k be the set of remaining objects. Note that $X^k = \mu^E(N^k)$.

The algorithm terminates when every agent is assigned an object. Since at least one agent receives an object and is removed at each step, the algorithm is over in at most n steps. Let μ^{TTC} denote the resulting assignment.

Note that at the beginning of each step k where the algorithm is not yet over, the set of available agents N^{k-1} is a nonempty finite set. Furthermore, the directed graph that we define over N^{k-1} is such that every agent in N^{k-1} points to an agent in N^{k-1}. Therefore, the existence of a cycle is guaranteed by question 4 in problem set 1.

Let's make a clarification analogous to what we noted after we described the DA algorithm. The TTC algorithm is used to describe the assignment μ^{TTC}. We are not trying to suggest that in real life, the agents in the housing market are indeed going through a dynamic process involving cyclic exchanges described by the algorithm. The algorithm is just a method to compute the assignment μ^{TTC}, and not an actual physical process.

Example 6 Consider the housing market given by $N = \{1, 2, 3, 4, 5, 6\}$, $X = \{a, b, c, d, e, f\}$,

R_1	R_2	R_3	R_4	R_5	R_6
a	a	b	c	d	c
$\vdots$	$\underline{d}$	$\vdots$	$\vdots$	f	e
$\underline{b}$	$\vdots$	$\underline{c}$	$\underline{a}$	$\underline{e}$	$\underline{f}$
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$\mu^E = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ b & d & c & a & e & f \end{pmatrix}$$

Consider the TTC algorithm for the above housing market:

Step 1: The only cycle formed at this step is: 1 points to 4, 4 points to 3, and 3 points to 1. Therefore, 1 receives a, 4 receives c, and 3 receives b. After we remove these agents and objects, the remaining agents and objects are $N^1 = \{2, 5, 6\}$ and $X^1 = \{d, e, f\}$.

Step 2: The only cycle formed at this step is the self-cycle: 2 points to 2. Therefore, 2 receives d. After we remove 2 and d, the remaining agents and objects are $N^2 = \{5, 6\}$ and $X^2 = \{e, f\}$.

Step 3: Agent 5 points to 6 and 6 points to 5. Therefore, 5 receives f and 6 receives e. There are no remaining agents and objects, i.e., $N^3 = X^3 = \emptyset$, so the algorithm terminates.

The resulting assignment is given by:

$$\mu^{TTC} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ a & d & b & c & f & e \end{pmatrix}$$

The next result shows that the outcome of the TTC algorithm, give us the unique core allocation of a housing market.

Theorem 3 *(Roth & Postlewaite (1977)) In a housing market* (N, X, R, μ^E)*, the outcome* μ^{TTC} *of the TTC algorithm is the unique assignment in the core.*

Proof: Covered in class. $\square$

4.1 Prices Supporting μ^{TTC} as a Walrasian Equilibrium

An important theorem in general equilibrium theory states that every Walrasian equilibrium allocation is in the core. This theorem holds in the special case of the housing market. We will next prove that a converse of this result also holds in the housing market: There is a price vector p such that (μ^{TTC}, p) is a Walrasian equilibrium.[1] Let's first define what we mean by a Walrasian equilibrium of a housing market.

Definition 19 Given a housing market (N, X, R, μ^E), a vector of prices $p = (p_x)_{x\in X} \in \mathbb{R}^n_+$ and an assignment μ constitute a **Walrasian equilibrium** (μ, p) if for all $i \in N$:

1. $p_{\mu(i)} \leq p_{\mu^E(i)}$, and
2. $\forall x \in X : p_x \leq p_{\mu^E(i)}$ implies $\mu(i) R_i x$.

In the above definition, we interpret the vector p as the prices of the objects in the market, and the assignment μ as the optimal consumption choices of the agents under the prices p and endowments defined by μ^E. Condition 1 states that the object that i consumes at μ is affordable for i, in the sense that its price does not exceed the price of the house that i owns. Condition 2 states that $\mu(i)$ is i's top choice among all objects that he can afford. Therefore, if the prices of the objects are given by p, then the wealth of agent $i \in N$ is $p_{\mu^E(i)}$, and he optimally demands the object $\mu(i)$.

The following result shows that the unique core allocation of a housing market given by the TTC outcome, can be made into a Walrasian equilibrium along with some price vector p. The proof shows how to construct the supporting price vector p from the TTC algorithm.

Proposition 5 *(Shapley & Scarf (1974)) For any housing market* (N, X, R, μ^E)*, there is a price vector* p *such that* (μ^{TTC}, p) *is a Walrasian equilibrium.*

Proof: It is easy to see that if we choose the supporting prices in the reverse order of the step in which an object is assigned in the TTC algorithm: e.g. $p_x = n - k$ if $x \in X^{k-1} \setminus X^k$; then (μ^{TTC}, p) is a Walrasian equilibrium. $\square$

[1] This result does not hold beyond the housing market model. That is, there are general equilibrium models where not all core allocations can be made into Walrasian equilibria.

4.2 Strategic Aspects and Characterization of the Core SCF

Let's start by defining social choice functions in the context of a housing market. In the following, fix (N, X, μ^E). Let $\mathcal{A}$ denote the set of all assignments and let $\mathcal{R}$ denote the set of all linear orders over the objects X.

Definition 20 A **social choice function (SCF)** is a function $f : \mathcal{R}^n \to \mathcal{A}$. For every preference profile $R \in \mathcal{R}^n$, and every agent $i \in N$, let $f_i(R)$ denote the object allocated to i at the assignment $f(R)$.

Definition 21 A SCF $f : \mathcal{R}^n \to \mathcal{A}$ is **strategyproof** if for every $R \in \mathcal{R}^n$, $i \in N$, and $R'_i \in \mathcal{R}$:

$$f_i(R_i, R_{-i})\, R_i\, f_i(R'_i, R_{-i})$$

In the above equation, $R_{-i} \in \mathcal{R}^{n-1}$ denotes the vector of preference profiles of all the agents other than i, at the preference profile R.

Let f^{TTC} denote the social choice function that associates with each preference profile $R \in \mathcal{R}^N$, the unique core allocation of the housing market (N, X, R, μ^E) given by the TTC algorithm.

We next show that f^{TTC}, equivalently the core SCF, is strategy proof.

Proposition 6 *Given any (N, X, μ^E), f^{TTC} is strategyproof.*

Proof: Covered in class. □

We already noted that assignments in the core are Pareto efficient and individually rational. Therefore, along with the above result this shows that f^{TTC}, equivalently the core SCF, is strategyproof, Pareto efficient, and individually rational. We next present an interesting characterization by Ma (1994), showing that the only SCF that satisfies these three properties is f^{TTC}.

Theorem 4 *(Ma (1994)) Fix (N, X, μ^E). Then, a SCF f is individually rational, strategyproof, and Pareto efficient if and only if $f = f^{TTC}$.*

Proof: We already argued the "$\Leftarrow$" part. The "$\Rightarrow$" part is omitted. □

5 The College Admissions Model

The college admissions model is again a two-sided matching model, in which the two sides of the market consist of colleges and students. In the model, colleges are allowed to enroll multiple students whereas each student can attend at most one college. So the college admissions model is called a *many-to-one* matching model. Most real-life applications of two-sided matching markets are many-to-one, since members from one side (schools, firms, hospitals) can be matched to more than one member from the other side (students, workers, interns). Therefore, it is important to know how we can extend our approach in Section 2 to the many-to-one model. Below we will describe the college admissions model, matchings, stability, and show how one can extend the DA algorithms to find stable matchings. Most results from Section 2 extend to the many-to one case.

Definition 22 A **college admissions** model (C, S, R, q) is a tuple consisting of disjoint finite sets of colleges $C = \{c_1, \dots, c_n\}$, and students $S = \{s_1, \dots, s_p\}$, a vector $q = (q_c)_{c\in C} = (q_{c_1}, \dots, q_{c_n})$ where q_c denotes the quota/capacity/number of available seats at college c, and a preference profile $R = ((R_c)_{c\in C}, (R_s)_{s\in S}) = (R_{c_1}, \dots, R_{c_n}, R_{s_1}, \dots, R_{s_p})$, such that:

(a) $\forall c \in C : R_c$ is a linear order over $S \cup \{c\}$, and

(b) $\forall s \in S : R_s$ is a linear order over $C \cup \{s\}$.

The college admissions model above specifies colleges' preferences over individual students. We will not explicitly model colleges' preferences over sets of students, i.e., over incoming classes. Our formulation of stability will implicitly rely on the assumption that from the perspective of a college, there is no substitution nor complementarity among students, i.e., a college's ranking over students is not affected by the other students in the class. Note also that if $q = (1, \dots, 1)$, then the college admissions model reduces to the marriage market model.

We next define matchings in a college admissions model.

Definition 23 Given a college admissions model (C, S, R, q), a **matching** is a function $\mu : S \to C$ such that:

1. $\forall s \in S$, $\mu(s) \in C \cup \{s\}$
2. $\forall c \in C$, $|\mu^{-1}(c)| \leq q_c$

The first condition above requires that every student is either matched to a college or remains unmatched. Note that the set $\mu^{-1}(c) = \{s \in S : \mu(s) = c\}$ is the set of subset of students matched to college c under the matching μ. Therefore, the second condition above is a feasibility requirement, the number of students matched to each college can not exceed the capacity of the college.

We next extend the definition of stability to the college admissions model.

Definition 24 Given a college admissions model (C, S, R, q), a matching μ is **stable** if it is

1. **Individually Rational (IR)**:
   (a) $\mu(s) R_s s$ for all $s \in S$.
   (b) $s R_c c$ for all $c \in C$ and $s \in \mu^{-1}(c)$.

2. **Pairwise Stable**: There is no pair $(s, c) \in S \times C$ such that $c P_s \mu(s)$ and one of the following holds:
   (a) $s P_c s'$ for some $s' \in \mu^{-1}(c)$, or
   (b) $s P_c c$ and $|\mu^{-1}(c)| < q_c$.

The IR condition again requires that (a) students can only be matched to acceptable colleges, and (b) colleges can only be matched to acceptable students. We have to be a little more careful in extending pairwise stability to the many-to-one model, because now when a student-college pair (s, c) form a blocking pair if the student strictly prefers the college over her current match $\mu(s)$ (possibly herself), and either (a) college c is matched to a student s' that is strictly worse than s for c, or (b) college c has some empty seats and finds student s acceptable.

Given a college admissions model (C, S, R, q), we can define the student-proposing DA algorithm, by modifying the earlier algorithm to allow for college c to have a waiting list with q_c slots instead of one slot. We can also define the college-proposing DA algorithm, where in the first step, each college c proposes to its top (at most) q_c acceptable students concurrently, and in every subsequent step, makes (at most) as many new proposals to acceptable students as the rejections it received in the the previous step. Let μ_S denote the outcome of the student-proposing DA algorithm, and let μ_C be the outcome of the college-proposing DA algorithm. Then, it can be similarly shown that μ_S is the best stable matching for the students, and μ_C is the best stable matching for the hospitals. It is also possible to extend Theorem 2 about incentives to show that the student-optimal stable mechanism is strategyproof for students.

6 Optional Additional Reading

The original paper by Gale and Shapley, and Chapters 1–5 of **RS** are related directly to the material covered in these lecture notes.

- Gale and Shapley (1962), "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69(1), 9–15.

RS Roth and Sotomayor, *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*, Cambridge University Press, 1990.

If you are interested in knowing more about the actual real-life applications, Al Roth has a number of surveys in his website, e.g.:

- Roth (2008), "What Have we Learned from Market Design?," *The Economic Journal*, 118, 285–310.

Also, the following recent paper introduces an interesting generalization of the DA algorithm where agents from both sides may propose:

- Dworczak (2015), "Deferred Acceptance with Compensation Chains," Stanford University, mimeo.