Mathematical Economics Seminar I

Two-Sided Matching Theory[1]

Myeonghwan Cho[2]

June 13, 2016

---

1 One-to-one matching: the marriage model

1.1 Basic model

- A *marriage market* is a triple (M, W, P):
  - M is a finite set of men.
  - W is a finite set of women.
  - P is a preference profile.
    $P(m)$ is a preference relation of man m over $W \cup \{m\}$.
    $P(w)$ is a preference relation of woman w over $M \cup \{w\}$.
- Representation of preference:

$$P(m) = w_1, w_5, [w_2, w_3], m, w_4 \tag{1.1}$$

  or equivalently

$$w_1 \succ_m w_5 \succ_m w_2 \sim_m w_3 \succ_m m \succ_m w_4 \tag{1.2}$$

- If agent $k \in M \cup W$ prefers to remain single rather than be matched to agent j, i.e., if $k \succ_k j$, then j is *unacceptable* to k.
- Sometimes, we ignore the preference of men (women) on the unacceptable agents because they are not matched with unacceptable women (men) in plausible matchings.

[1] This note is prepared for the class at University of Seoul. This note follows Roth and Sotomayor (1990) and theorems are numbered as in the book. This note may have some typos and errors. If you find any error or typo, please let me know via e-mail (chomhmh@uos.ac.kr).

[2] Department of Economics, University of Seoul. E-mail: chomhmh@skku.edu

This allows us to consider the preference of each man (woman) over the acceptable women (men) and to represent his (her) preference as

$$P(m) = w_1, w_5, [w_2, w_3]$$

or equivalently

$$w_1 \succ_m w_5 \succ_m w_2 \sim_m w_3$$

instead of (1.1) and (1.2).

- An outcome of the marriage market (M, W, P) is a *matching*

$$\mu : M \cup W \to M \cup W$$

such that $\mu(m) = w$ if and only if $\mu(w) = m$ and for all m and for all w, $\mu(w) \in M \cup \{w\}$ and $\mu(m) \in W \cup \{m\}$.

- Representation of matching: $M = \{m_1, m_2, m_3, m_4, m_5\}$ and $W = \{w_1, w_2, w_3, w_4\}$

$$\mu = \begin{array}{ccccc} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{array}$$

or equivalently

$$\mu = \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & (m_5) \\ m_2 & m_3 & m_4 & m_1 & m_5 \end{array}$$

- Given the preference $\succsim_m$ of $m \in M$ (or $w \in W$), we can induce the preference of m (or $w \in W$) over the matchings: for matchings μ and μ',

$$\mu \succsim_m \mu' \iff \mu(m) \succsim_m \mu'(m).$$

1.2 Stable matchings

- A matching μ is *blocked by an individual* $k \in M \cup W$, if k prefers being single to being matched with $\mu(k)$, i.e., $k \succ_k \mu(k)$.

- A matching μ is *blocked by a pair* $(m, w) \in M \times W$, if they prefer each other to their partners under μ, i.e.,

$$w \succ_m \mu(m) \text{ and } m \succ_w \mu(w).$$

- A matching μ is *stable*, if it is not blocked by any individual or pair of agents.

- When the men and women have strict preference, a stable matching is Pareto-efficient. (Why?)

- The *weak core* is the set of matchings μ such that there exists no matching ν and coalition $T \subseteq M \cup W$ such that for all $k \in T$, $\nu(k) \succ_k \mu(k)$ and $\nu(k) \in T$.

- The *(strong) core* is the set of matchings μ such that there exists no matching ν and coalition $T \subseteq M \cup W$ such that for all $k \in T$, $\nu(k) \succsim_k \mu(k)$ with $\nu(i) \succ_i \mu(i)$ for some $i \in T$, and for all $k \in T$ in $\nu(k) \in T$.

- $\mathcal{S}$: set of stable matchings

  $\mathcal{C}^W$: set of stable matchings

  $\mathcal{C}^S$: set of stable matchings

Lemma 1. *In the marriage model, when preferences are general,*

$$\mathcal{C}^S \subseteq \mathcal{C}^W = S.$$

When preferences are strict,

$$\mathcal{C}^S = \mathcal{C}^W = S.$$

- A stable matching is also referred to as a *core matching.*

Men-proposing deferred acceptance algorithm

Step 0: If some preferences are not strict, arbitrarily break ties (e.g. if some m is indifferent between w and w', order them consecutively in an arbitrary manner)

Step 1: Each man m proposes to his 1st choice (if he has any acceptable choices).
Each woman rejects any unacceptable proposals and, if more than one acceptable proposal is received, "holds" the most preferred.
If no proposals are rejected, then match each woman to the man (if any) whose proposal she

is “holding” and terminate the procedure.

$\vdots$

Step k: Any man who was rejected at step $k-1$ makes a new proposal to his most preferred acceptable mate who has not yet rejected him. (If no acceptable choices remain, he makes no proposal.)
Each woman “holds” her most preferred acceptable offer to date, and rejects the rest.
If no proposals are rejected, then match each woman to the man (if any) whose proposal she is “holding” and terminate the procedure.

Example 1 (Example of the deferred acceptance procedure)**.** Consider a marriage market (M, W, P) such that

$$\begin{array}{ll}
P(m_1) = w_1, w_2, w_3, w_4 & \qquad P(w_1) = m_2, m_3, m_1, m_4, m_5 \\
P(m_2) = w_4, w_2, w_3, w_1 & \qquad P(w_2) = m_3, m_1, m_2, m_4, m_5 \\
P(m_3) = w_4, w_3, w_1, w_2 & \qquad P(w_3) = m_5, m_4, m_1, m_2, m_3 \\
P(m_4) = w_1, w_4, w_3, w_2 & \qquad P(w_4) = m_1, m_4, m_5, m_2, m_3 \\
P(m_5) = w_1, w_2, w_4 &
\end{array}$$

The deferred acceptance procedure in which the men make an offer results in the stable matching

$$\mu_M = \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{array}.$$

The deferred acceptance procedure in which the women make an offer results in the stable matching

$$\mu_W = \begin{array}{ccccc} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{array}.$$

$\diamondsuit$

Theorem 1 (Gale and Shapley)**.** *A stable matching exists for every marriage market.*

Proof. Consider the men-proposing deferred acceptance algorithm.

The algorithm stops after any step in which no man is rejected. At this point, every man is either engaged to some woman or has been rejected by every woman on his list of acceptable women. The marriages are now consummated, with each man being matched to the woman to whom he is engaged. Women who did not receive any acceptable proposal, and men who were rejected by all women acceptable to them, will stay single.

This completes the description of the algorithm, except that we have described it as if all agents have strict preferences. The modification required in case some man or woman is indifferent between two or more possible mates is simple. At any step of the algorithm at which some agent must indicate a choice between two mates who are equally well liked, introduce some fixed "tie-breaking" rule (e.g., when an agent is indifferent, proceed as if the preferences are according to alphabetical order of family names, or as if agents prefer mates who are closer to them in age, etc.). Such a tie-breaking rule therefore specifies, arbitrarily, to which woman a man will propose when he is indifferent about his next proposal, and which man a woman will keep engaged when she is indifferent among more than one most favored suitors.

The algorithm must eventually stop because there are only a finite number of men and women, and no man proposes more than once to any woman. The outcome that it produces is a matching, since each man is engaged at any step to at most one woman, and each woman is engaged at any step to at most one man. Furthermore, this matching is individually rational, since no man or woman is ever engaged to an unacceptable partner.

To see that the matching μ produced by the algorithm is stable, suppose some man m and woman w are not married to each other at μ, but m prefers w to his own mate at μ. Then woman w must be acceptable to man m, and so he must have proposed to her before proposing to his current mate (or before being rejected by all of the women he finds acceptable). Since he was not engaged to w when the algorithm stopped, he must have been rejected by her in favor of someone she liked at least as well. Therefore w is matched at μ to a man whom she likes at least as well as man m, since preferences are transitive (and hence acyclic), and so m and w do not block the matching μ. Since the matching is not blocked by any individual or by any pair, it is stable. $\square$

Example 2 (Roommate problem (Gale and Shapley))**.** There is a single set of n people who can be matched in pairs (to be roommates in a college dormitory, or partners in paddling a canoe). Each person in the set ranks the $n-1$ others in order of preference. An outcome is a matching, which is a partition of the people into pairs. To keep things simple, suppose the number n of people is even. A stable matching is a matching such that no two persons who are not roommates both prefer each other to their actual partners.

Consider four people: a, b, c, and d, with the following preferences:

$$
\begin{aligned}
P(a) &= b, c, d \\
P(b) &= c, a, d \\
P(c) &= a, b, d \\
P(d) &= \text{arbitrary}
\end{aligned}
$$

In this example, we can see that there is no stable matching. $\diamondsuit$

Example 3 (Man-woman-child marriage problem (Alkan))**.** There are three sets of people: men, women, and children. A matching is a division of the people into groups of three, containing one man, one woman, and one child. Each person has preferences over the sets of pairs he or she might possibly be matched with. A man, woman, and child (m, w, c) *block* a matching μ if m prefers (w, c) to $\mu(m)$; w prefers (m, c) to $\mu(w)$, and c prefers (m, w) to $\mu(c)$. A matching is stable only if it is not blocked by any such three agents.

Consider three men, three women, and three children, with the following preferences:

$$
\begin{aligned}
P(m_1) &= (w_1, c_3), (w_2, c_3), (w_1, c_1), \ldots \text{(arbitrary)} \\
P(m_2) &= (w_2, c_3), (w_2, c_2), (w_3, c_3), \ldots \text{(arbitrary)} \\
P(m_3) &= (w_3, c_3), \ldots \text{(arbitrary)}
\end{aligned}
$$

$$
\begin{aligned}
P(w_1) &= (m_1, c_1), \ldots \text{(arbitrary)} \\
P(w_2) &= (m_2, c_3), (m_1, c_3), (m_2, c_2), \ldots \text{(arbitrary)} \\
P(w_3) &= (m_2, c_3), (m_3, c_3), \ldots \text{(arbitrary)}
\end{aligned}
$$

$$
\begin{aligned}
P(c_1) &= (m_1, w_1), \ldots \text{(arbitrary)} \\
P(c_2) &= (m_2, w_2), \ldots \text{(arbitrary)} \\
P(c_3) &= (m_1, w_3), (m_2, w_3), (m_1, w_2), (m_3, w_3), \ldots \text{(arbitrary)}
\end{aligned}
$$

1. All matchings that give m_1 (*resp.* m_2 and w_2) a better family than (m_1, w_1, c_1) (*resp.* (m_2, w_2, c_2)) are unstable: Any matching containing (m_1, w_1, c_3) or (m_2, w_2, c_3) is blocked by (m_3, w_3, c_3); Any matching containing (m_1, w_2, c_3) is blocked by (m_2, w_3, c_3).

2. Any matching that does not contain (m_1, w_1, c_1) (*resp.* (m_2, w_2, c_2)) is either blocked by (m_1, w_1, c_1) (*resp.* (m_2, w_2, c_2)) or is unstable as already shown in 1 above.

3. (m_1, w_2, c_3) blocks any matching that contains (m_1, w_1, c_1) and (m_2, w_2, c_2).

Thus, there is no stable matching in this example. $\diamondsuit$

Example 4 (Many-to-one matching)**.** Consider a set of firms and a set of workers. Each worker can work for at most one firm and has preferences over those firms he or she is willing to work for. Each firm can hire as many workers as it wishes and has preferences over those subsets of workers it is willing to employ. It is clear what a matching is in this case, and a firm F and a subset of workers C block a matching μ if F prefers C to the set of workers assigned to it at μ, and every worker in C who is not assigned to F prefers F to the firm he or she is assigned by μ. Consider two firms and three workers with the following preferences:

$$
\begin{aligned}
P(F_1) &= \{w_1, w_3\}, \{w_1, w_2\}, \{w_2, w_3\}, \{w_1\}, \{w_2\} \\
P(F_2) &= \{w_1, w_3\}, \{w_2, w_3\}, \{w_1, w_2\}, \{w_3\}, \{w_1\}, \{w_2\} \\
P(w_1) &= F_2, F_1 \\
P(w_2) &= F_2, F_1 \\
P(w_3) &= F_1, F_2
\end{aligned}
$$

The only individually rational matchings without unemployment are:

$$
\begin{aligned}
\mu_1 &= \begin{matrix} F_1 & F_2 \\ \{w_1, w_3\} & \{w_2\} \end{matrix}, && \text{which is blocked by } (F_2, w_1) \\
\mu_2 &= \begin{matrix} F_1 & F_2 \\ \{w_1, w_2\} & \{w_3\} \end{matrix}, && \text{which is blocked by } (F_2, \{w_1, w_3\}) \\
\mu_3 &= \begin{matrix} F_1 & F_2 \\ \{w_2, w_3\} & \{w_1\} \end{matrix}, && \text{which is blocked by } (F_2, \{w_1, w_2\}) \\
\mu_4 &= \begin{matrix} F_1 & F_2 \\ \{w_2\} & \{w_1, w_3\} \end{matrix}, && \text{which is blocked by } (F_1, \{w_2, w_3\}) \\
\mu_5 &= \begin{matrix} F_1 & F_2 \\ \{w_1\} & \{w_2, w_3\} \end{matrix}, && \text{which is blocked by } (F_2, \{w_1, w_3\})
\end{aligned}
$$

Now observe that any matching that leaves w_1 unmatched is blocked either by (F_1, w_1) or by (F_2, w_1); any matching that leaves w_2 unmatched is blocked either by (F_1, w_2), (F_2, w_2), or $(F_2, \{w2, w3\})$. Finally, any matching that leaves w_3 unmatched is blocked by $(F_2, \{w_1, w_3\})$. $\diamondsuit$

Theorem 2 (Gale and Shapley)**.** *When all men and women have strict preferences, there always exists a men-optimal stable matching (that every man likes at least as well as any other stable matching), and a women-optimal stable matching.*

Furthermore, the matching μ_M produced by the deferred acceptance algorithm with men-proposing is the men-optimal stable matching. The women-optimal stable matching is the matching μ_W produced by the algorithm when the women propose.

- A woman w and a man m are *achievable* for each other in a marriage market (M, W, P) if m and w are paired at some stable matching.

Proof of Theorem 2. When all men and women have strict preferences, we will show that in the deferred acceptance algorithm with men proposing, no man is ever rejected by an achievable woman. Consequently the stable matching μ_M that is produced matches each man to his most preferred achievable woman, and is therefore the (unique) men-optimal stable matching.

The proof is by induction. Assume that up to a given step in the procedure no man has yet been rejected by a woman who is achievable for him. At this step, suppose woman w rejects man m. If she rejects m as unacceptable, then she is unachievable for him, and we are done. If she rejects m in favor of man m', whom she keeps engaged, then she prefers m' to m. We must show that w is not achievable for m.

We know m' prefers w to any woman except for those who have previously rejected him, and hence (by the inductive assumption) are unachievable for him. Consider a hypothetical matching μ that matches m to w and everyone else to an achievable mate. Then m' prefers w to his mate at μ. So the matching μ is unstable, since it is blocked by m' and w, who each prefer the other to their mate at μ. Therefore there is no stable matching that matches m and w, and so they are unachievable for each other, which completes the proof. □

- Let $\mu \succ_M \mu'$ denote that all men like μ at least as well as μ', with at least one man having strict preference.

  $\mu \succsim_M \mu'$ means that $\mu \succ_M \mu'$ or that all men are indifferent between μ and μ'.

- Then $\succ_M$ is a partial order on the set of matchings, representing the common preferences of the men.

- Similarly, we define $\succ_W$ and $\succsim_W$ as the common preference of the women.

Theorem 3 (Knuth)**.** *When all agents have strict preferences, the common preferences of the two sides of the market are opposed on the set of stable matchings: if μ and μ' are stable matchings, then all men like μ at least as well as μ' if and only if all women like μ at least as well as μ'. That is, $\mu \succ_M \mu'$ if and only if $\mu' \succ_W \mu$.*

Proof. Let μ and μ' be stable matchings such that $\mu \succ_M \mu'$. We will show that $\mu' \succ_W \mu$.

Suppose it is not true that $\mu' \succ_W \mu$. Then there must be some woman w who prefers μ to μ'. Then, woman w has a different mate at μ and μ', and consequently so does man $m = \mu(w)$. Since m also has strict preferences, m and w form a blocking pair for the matching μ'. This contradicts the assumption that μ' is stable. Therefore, $\mu' \succ_W \mu$, as required. □

Corollary 1. *When all agents have strict preferences, the M-optimal stable matching is the worst stable matching for the women; that is, it matches each woman with her least preferred achievable mate. Similarly, the W-optimal stable matching matches each man with his least preferred achievable mate.*

- For any matchings μ and μ', define a function

  $$\lambda = \mu \vee_M \mu'$$

  on the set $M \cup W$ as, for all $m \in M$,

  $$\lambda(m) = \begin{cases} \mu(m) & \text{if } \mu(m) \succ_m \mu'(m) \\ \mu'(m) & \text{otherwise} \end{cases}$$

  and for all $w \in W$,

  $$\lambda(w) = \begin{cases} \mu(m) & \text{if } \mu(m) \prec_w \mu'(m) \\ \mu'(m) & \text{otherwise} \end{cases}$$

- For any matchings μ and μ', define a function

  $$\nu = \mu \wedge_M \mu'$$

  on the set $M \cup W$ as, for all $m \in M$,

  $$\nu(m) = \begin{cases} \mu(m) & \text{if } \mu(m) \prec_m \mu'(m) \\ \mu'(m) & \text{otherwise} \end{cases}$$

  and for all $w \in W$,

  $$\nu(w) = \begin{cases} \mu(m) & \text{if } \mu(m) \succ_w \mu'(m) \\ \mu'(m) & \text{otherwise} \end{cases}$$

- The functions $\lambda = \mu \vee_M \mu'$ and $\nu = \mu \wedge_M \mu'$ are not necessarily matchings, and even when λ and ν are matchings, they are not necessarily stable.

Lattice

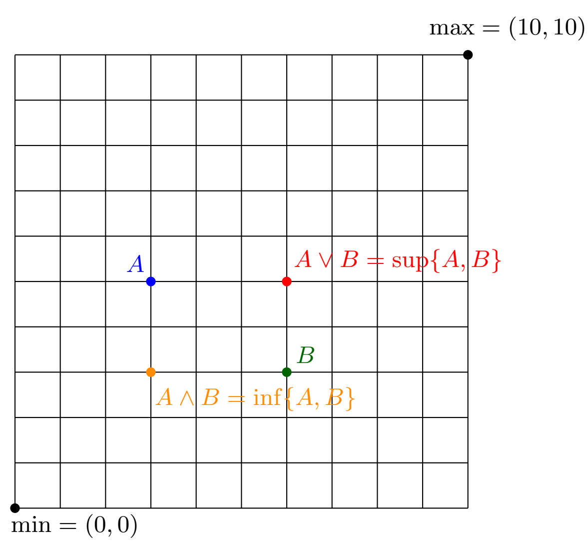
- L: a set endowed with a partial order $\succeq$.
- An *upper bound* of a subset $X \subset L$ is an element $a \in L$ such that $a \succeq x$ for all $x \in X$.

  An *lower bound* of a subset $X \subset L$ is an element $a \in L$ such that $a \preceq x$ for all $x \in X$.
- $\sup X$: *supremum* of X, the least upper bound of X if it exists.

  $\inf X$: *infimum* of X, the greatest lower bound of X if it exists.
- Note that, if $\sup X$ (or, $\inf X$) exists, it is unique.
- A *lattice* is a partially ordered set L, if any two of whose elements x and y have a "sup", denoted by $x \vee y$ and an "inf", denoted by $x \wedge y$.
- $x \vee y$ is called a *join* of x and y.

  $x \wedge y$ is called a *meet* of x and y.
- A lattice is *complete* when each of its subsets X has a "sup" and an "inf" in L.

Example 5. $\{0, 1, \ldots, 10\}^2$ and usual vector partial order $\geq$



Theorem 4 (Lattice theorem (Conway))**.** *When all preferences are strict, if μ and μ' are stable matchings, then the functions $\lambda = \mu \vee_M \mu'$ and $\nu = \mu \wedge_M \mu'$ are both matchings and they are both stable.*

Proof. Step 1: λ is a matching ($\lambda(m) = w$ if and only if $\lambda(w) = m$).
$[\lambda(m) = w \Rightarrow \lambda(w) = m]$ By stability of μ and μ', this holds.

$[\lambda(w) = m \Rightarrow \lambda(m) = w]$ Let $M' = \{m : \lambda(m) \in W\} = \{m : \mu(m) \in W \text{ or } \mu'(m) \in W\}$. By the only if direction, $\lambda(M')$ is contained in $\{w : \lambda(w) \in M\}$ which (by definition of λ) equals $W' = \{w : \mu(w) \in M \text{ and } \mu'(w) \in M\}$, which is the same size as $\mu(W')$. But $\lambda(M')$ is the same size as M' (since $\lambda(m) = \lambda(m') = w$ only if $m = m' = \lambda(w)$), which is at least as large as $\mu(W')$, so $\lambda(M')$ and W' are the same size and $\lambda(M') = W'$. Hence for $w \in W'$, $\lambda(w) = m$ for some $m \in M'$, so $\lambda(w) = m$. If w is not in W' then $\lambda(w) = w$. So if $\lambda(w) = m$ then $\lambda(m) = w$.

From a symmetric argument, ν is also a matching.

Step 2: λ is stable.
Suppose (m, w) blocks λ. Then $w \succ_m \lambda(m)$ and $m \succ_w \lambda(w)$. $w \succ_m \lambda(m)$ implies that $w \succ_m \mu(m)$ and $w \succ_w \mu'(m)$. If $\lambda(w) = \mu(w)$, then (m, w) blocks μ (contradiction!). If $\lambda(w) = \mu'(w)$, then (m, w) blocks μ' (contradiction!). This shows that λ is stable.

By the symmetric argument ν is also stable. □

- We can think of λ as asking men to point to their preferred mate from two stable matchings, and asking women to point to their less preferred mate.
- Theorem 4 says that:
  - No two men point to the same woman.
  - Every woman points back at the man pointing to her.
  - The resulting matching is stable.

Example 6 (The lattice of stable matchings (Knuth))**.**

$$\begin{aligned} P(m_1) &= w_1, w_2, w_3, w_4 & \qquad P(w_1) &= m_4, m_3, m_2, m_1 \\ P(m_2) &= w_2, w_1, w_4, w_3 & P(w_2) &= m_3, m_4, m_1, m_2 \\ P(m_3) &= w_3, w_4, w_1, w_2 & P(w_3) &= m_2, m_1, m_4, m_3 \\ P(m_4) &= w_4, w_3, w_2, w_1 & P(w_4) &= m_1, m_2, m_3, m_4 \end{aligned}$$

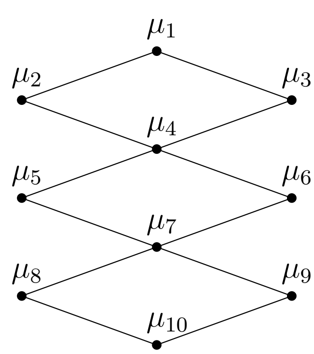
There are then stable matchings where w_1, w_2, w_3 and w_4 are matched respectively to

$$
\begin{array}{lcl}
\mu_1 & : & m_1\ m_2\ m_3\ m_4 \\
\mu_2 & : & m_2\ m_1\ m_3\ m_4 \\
\mu_3 & : & m_1\ m_2\ m_4\ m_3 \\
\mu_4 & : & m_2\ m_1\ m_4\ m_3 \\
\mu_5 & : & m_3\ m_1\ m_4\ m_2 \\
\mu_6 & : & m_2\ m_4\ m_1\ m_3 \\
\mu_7 & : & m_3\ m_4\ m_1\ m_2 \\
\mu_8 & : & m_4\ m_3\ m_1\ m_2 \\
\mu_9 & : & m_3\ m_4\ m_2\ m_1 \\
\mu_{10} & : & m_4\ m_3\ m_2\ m_1
\end{array}
$$

We can see that

$$
\begin{array}{ll}
\mu_2 \wedge_M \mu_3 = \mu_4 & \qquad \mu_2 \vee_M \mu_3 = \mu_1 \\
\mu_5 \wedge_M \mu_6 = \mu_7 & \qquad \mu_5 \vee_M \mu_6 = \mu_4 \\
\mu_8 \wedge_M \mu_9 = \mu_{10} & \qquad \mu_8 \vee_M \mu_9 = \mu_7
\end{array}
$$

The lattice structure with the stable matchings can be represented as follows:



$\diamondsuit$

Lemma 2 (Decomposition lemma (Knuth))**.** *Let μ and μ' be stable matchings in $(M, W, \succ)$, where all preferences are strict. Let $M(\mu)$ be the set of men m such that $\mu(m) \succ_m \mu'(m)$ and $W(\mu)$ be the set of women w such that $\mu(w) \succ_w \mu'(w)$. Analogously define $M(\mu')$ and $W(\mu')$. Then μ and μ' map $M(\mu')$ onto $W(\mu)$ and $M(\mu)$ onto $W(\mu')$.*

Proof. Suppose $m \in M(\mu')$. Since $\mu'(m) \succ_m \mu(m) \succsim_m m$, we have $\mu'(m) \in W$. Let $w =$

$\mu'(m)$. If $\mu'(w) \succ_w \mu(w)$, (m, w) blocks μ (contradiction!). Thus, since the preferences are strict, $\mu(w) \succ_w \mu'(w)$ and so $w = \mu'(m) \in W(\mu)$. This implies $\mu'(M(\mu')) \subset W(\mu)$.

On the other hand, suppose $w \in W(\mu)$. Since $\mu(w) \succ_w \mu'(w) \succsim_w w$, we have $\mu(w) \in M$. Let $m = \mu(w)$. If $\mu(m) \succ_m \mu'(m)$, (m, w) would block μ' (contradiction!). Thus, since the preferences are strict, $\mu'(m) \succ_m \mu(m)$ and so $m \in M(\mu')$. This implies $\mu(W(\mu)) \subset M(\mu')$.

Since μ and μ' are one-to-one and $M(\mu')$ and $W(\mu)$ are finite, the conclusion follows. □

Theorem 5. *In a market (M, W, P) with strict preferences, the set of people who are single is the same for all stable matchings.*

Proof. Suppose m was matched under μ' but not under μ. Then $m \in M(\mu')$. But, from Lemma 2, μ maps $W(\mu)$ onto $M(\mu')$, so m is also matched under μ. This is a contradiction. □

Theorem 6 (Weak Pareto optimality for the men)**.** *There is no individually rational matching μ (stable or not) such that $\mu \succ_m \mu_M$ for all m in M.*

Proof. If μ were such a matching it would match every man m to some woman w who had rejected him in the algorithm in favor of some other man m' (i.e., even though m was acceptable to w). Hence all of these women, $\mu(M)$, would have been matched under μ_M. That is, $\mu_M(\mu(M)) = M$. Hence all of M would have been matched under μ_M and $\mu_M(M) = \mu(M)$. But since all of M are matched under μ_M, any woman who gets a proposal in the last step of the algorithm at which proposals were issued has not rejected any acceptable man, that is, the algorithm stops as soon as every woman in $\mu_M(M)$ has an acceptable proposal. So such a woman must be single at μ (since every man prefers μ to μ_M), which contradicts the fact that $\mu_M(M) = \mu(M)$. □

Example 7. μ_M is not in general (strongly) Pareto efficient for men.

Consider a marriage market (M, W, P) with $M = \{m_1, m_2, m_3\}$, $W = \{w_1, w_2, w_3\}$, and

$$\begin{array}{ll} P(m_1) = w_2, w_1, w_3 & P(w_1) = m_1, m_2, m_3 \\ P(m_2) = w_1, w_2, w_3 & P(w_2) = m_3, m_1, m_2 \\ P(m_3) = w_1, w_2, w_3 & P(w_3) = m_1, m_2, m_3 \end{array}.$$

Then

$$\mu_M = \begin{matrix} w_1 & w_2 & w_3 \\ m_1 & m_3 & m_2 \end{matrix}.$$

Let

$$\mu = \begin{matrix} w_1 & w_2 & w_3 \\ m_3 & m_1 & m_2 \end{matrix}.$$

Then, $\mu \succ_M \mu_M$. $\diamondsuit$

1.3 Strategic behavior

- The deferred acceptance algorithm is decentralized in the sense that each man (woman) sequentially makes an offer to a woman (man) to be matched with and each woman decides whether to accept the offer or not.
- The same outcome can be obtained under the centralized matching mechanism in which the men and the women submit a list of their preferences to the social planner and the social planner decides the matching that is obtained under the deferred acceptance algorithm according to the submitted preferences. $\Rightarrow$ Revelation principle

The preference revelation mechanism: ϕ

- A clearing house fix an matching mechanism ϕ that assigns each preference profile Q to a matching $\phi(Q)$.

  Here, Q is a stated preference profile that is not necessarily equal to the true preference profile P.
- A matching mechanism ϕ is *Pareto optimal* (or *Pareto efficient*), if ϕ always yields a Pareto optimal matching for any preference profile P.
- A matching mechanism ϕ is a *stable matching mechanism*, if ϕ always yields a stable matching for any preference profile P.
- Consider an agent i whose true preference is P_i which is represented with $\succsim_i$. A strategy (a stated preference Q_i) for agent i is a *dominant strategy* if, for all Q_{-i},

  $$\phi(Q_i, Q_{-i}) \succsim_i \phi(Q'_i, Q_{-i})$$

  for all Q'_i (stated preferences).
- A matching mechanism ϕ is *strategyproof*, if it makes it a dominant strategy for each agent to state his/her true preferences.

Example 8 (A strategyproof and Pareto optimal matching)**.** Place the men in some order $m_1, m_2, \ldots, m_n$. Consider a matching mechanism ϕ that matches m_1 to his stated first choice, m_2 to his stated first choice of possible mates remaining after m_1's matched woman is removed from the market, and any m_k to his stated first choice after m_1's, m_2's, … m_k's matched women are removed from the market. Then, ϕ is strategyproof and Pareto optimal. But, ϕ is not a stable matching mechanism. $\diamondsuit$

Example 9. Consider a marriage market (M, W, P) with $M = \{m_1, m_2, m_3, m_4, m_5\}$, $W = \{w_1, w_2, w_3, w_4\}$, and

$$\begin{array}{ll} P(m_1) = w_1, w_2, w_3, w_4 & \qquad P(w_1) = m_2, m_3, m_1, m_4, m_5 \\ P(m_2) = w_4, w_2, w_3, w_1 & \qquad P(w_2) = m_3, m_1, m_2, m_4, m_5 \\ P(m_3) = w_4, w_3, w_1, w_2 & \qquad P(w_3) = m_5, m_4, m_1, m_2, m_3 \\ P(m_4) = w_1, w_4, w_3, w_2 & \qquad P(w_4) = m_1, m_4, m_5, m_2, m_3 \\ P(m_5) = w_1, w_2, w_4 & \end{array}$$

The M-optimal stable matching is

$$\mu_M = \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{array}.$$

Consider now the preference P' in which all agents except w_1 state their preferences as P but w_1 misrepresents her preferences by stating

$$P'(w_1) = m_2, m_3, m_4, m_5, m_1$$

Then, the resulting matching μ'_M is

$$\mu'_M = \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & (m_5) \\ m_3 & m_1 & m_2 & m_4 & m_5 \end{array}.$$

Note that w_1 becomes better off by misrepresenting P' for her preference. $\diamondsuit$

Theorem 7 (Impossibility Theorem, Roth)**.** *No stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.*

Proof. For the proof, one example for which no stable matching mechanism induces a dominant strategy is sufficient. Consider a marriage market (M, W, P) with $M = \{m_1, m_2\}$,

$W = \{w_1, w_2\}$, and

$$\begin{array}{ll} P(m_1) = w_1, w_2 & \qquad\qquad P(w_1) = m_2, m_1 \\ P(m_2) = w_2, w_1 & \qquad\qquad P(w_2) = m_1, m_2 \end{array}.$$

Then there are two stable matchings:

$$\mu_M = \begin{array}{cc} w_1 & w_2 \\ m_1 & m_2 \end{array} \qquad\qquad \mu_W = \begin{array}{cc} w_1 & w_2 \\ m_2 & m_1 \end{array}.$$

If $\phi(P) = \mu_M$, then w_1 has an incentive to state that m_2 is only acceptable man (stating $\succsim'_{w_1}$). Note that the stable matching mechanism ϕ should yield μ_W under the misrepresentation of w_1's preference. If $\phi(P) = \mu_W$, then m_1 has an incentive to state that w_1 is only acceptable woman. Note that the stable matching mechanism ϕ should yield μ_M under the misrepresentation of m_1's preference. $\square$

- One might think from the proof of Theorem 7 that situations in which some agent can profitably manipulate his preferences are rare. But the next proposition says that such situations are not rare.

Theorem 8. *When any stable mechanism is applied to a marriage market in which preferences are strict and there is more than one stable matching, then at least one agent can profitably misrepresent his or her preferences, assuming the others tell the truth. (This agent can misrepresent in such a way as to be matched to his or her most preferred achievable mate under the true preferences at every stable matching under the misstated preferences.)*

Proof. By assumption, $\mu_M \neq \mu_W$. Suppose that when all agents state their true preferences, the mechanism selects a stable matching $\mu \neq \mu_W$. Let w be any woman such that $\mu_W(W) \succ_w \mu(w)$. (So w is not single at μ_W.) Now let w misrepresent her preferences by removing from her stated preference list of acceptable men all men who rank below $\mu_W(w)$. Note that μ_W will still be stable under these new preferences.

Let μ' be the stable matching selected by the mechanism for these new preferences. By Theorem 5, w is not single at μ' and hence she is matched with someone she likes at least as well as $\mu_W(w)$. In fact, since μ' is also stable under the original preferences, it follows that $\mu'(w) = \mu_W(w)$. But w prefers $\mu_W(w)$ to $\mu(w)$ so she prefers any stable μ' under the new preferences to μ.

If the mechanism originally selects the matching μ_W, then the symmetric argument can be made for any man m who strictly prefers μ_M. This completes the proof. $\square$

Theorem 9 (Dubins and Freedman, Roth)**.** *The mechanism that yields the M-optimal stable matching (in terms of the stated preferences) makes it a dominant strategy for each man to state his true preferences. (Similarly, the mechanism that yields the W-optimal stable matching makes it a dominant strategy for every woman to state her true preferences.)*

Theorem 10 (Dubins and Freeman)**.** *Let P be the true preferences of the agents, and let $\bar{P}$ differ from P in that some coalition $\bar{M}$ of the men misstate their preferences. Then there is no matching μ, stable for $\bar{P}$, which is preferred to μ_M by all members of $\bar{M}$.*

2 Many-to-one matching: The College Admissions Model

- A *college admission market* $(\mathbf{S}, \mathbf{C}, q, P)$ consists of:
  - a finite set of students $\mathbf{S}$,
  - a finite set of colleges $\mathbf{C}$,
  - a profile of quota $q = (q_c)_{c \in \mathbf{C}}$, where $q_c > 0$ is a quota for each $c \in C$, and
  - a preference profile P.

    $P(c)$ is a preference relation college $c \in \mathbf{C}$ over $\mathbf{S} \cup \{c\}$.

    $P(s)$ is a preference relation student $s \in \mathbf{S}$ over $\mathbf{C} \cup \{s\}$.
- Otherwise noted, we assume that the colleges and the students have strict preferences.
- We represents the college's preference P_c and student's preference P_s similarly to the marriage market.

  We also use the notation $\succsim_c$ and $\succsim_s$ for the college c's preference and the student s's preference, respectively.
- Because the colleges enroll a group of students, we need to specify how college's preferences over the collection of the sets of students are related to their preferences over individual students.
- *Responsive preferences*: for any set of students $S \subset \mathbf{S}$ with $|S| < q_c$, and any students s and $s' \in \mathbf{S} \backslash S$,

- $S \cup \{s\} \succ_c S \cup \{s'\}$ iff $s \succ_c s'$, and
- $S \cup \{s\} \succ_c S$ iff s is acceptable to c.

- Responsiveness of preference implies that the students are independent to the college's preference.

- Note that, although a college has a complete preference over the set of students, the responsiveness of preference over the collection of the sets of students does not ensure the completeness.

  However, there is a preference over the collection of the sets of students that is responsive, complete, and transitive.

- Given the preference of college over the collection of the sets of students, we can define the preference of the college on the matchings:

$$\mu \succ_c \mu' \iff \mu(c) \succ_c \mu'(c)$$

- For a college admission market, a *matching* μ is a mapping from $\mathbf{C} \cup \mathbf{S}$ into the set of unordered families of elements of $\mathbf{C} \cup \mathbf{S}$ such that:

  - $|\mu(s)| = 1$ for all $s \in \mathbf{S}$ and $\mu(s) = s$ if $\mu(s) \notin c$;
  - $|\mu(c)| = q_c$ for all $c \in \mathbf{C}$ and, if $|\mu(c) \cap \mathbf{S}| < q_c$, then $\mu(c)$ contains $q_c - |\mu(c) \cap \mathbf{S}|$ copies of c;
  - $\mu(s) = c$ if and only if $s \in \mu(c)$.

- Representation of matching μ:

  $(\mathbf{S}, \mathbf{C}, q, P)$ with $\mathbf{S} = \{s_1, s_2, s_3, s_4, s_5\}$, $\mathbf{C} = \{c_1, c_2\}$, and $q = (3, 2)$;

$$\mu = \begin{array}{ccc} c_1 & c_2 & (s_4) \\ s_1 s_3 c_1 & s_2 s_5 & s_4 \end{array}.$$

- A matching μ is *individually irrational* if $\mu(s) = c$ for some student s and college c such that either the student is unacceptable to the college or the college is unacceptable to the student.

  An individually irrational matching is said to be *blocked* by the relevant individual.

- A matching μ is *blocked by a college-student pair* (c, s), if they prefer each other to the matching μ:
  - $s \succ_c s'$ for some $s' \in \mu(c)$ or $s \succ_c c$ if $|\mu(c) \cap \mathbf{S}| < q_c$;
  - $c \succ_s \mu(s)$.
- As in the marriage model, a matching is *(pairwise) stable*, if it is not blocked by any individual or pair of agents.
- Colleges enroll to multiple students, it might not be enough to concentrate only on pairwise stability.

  However, the assumption of responsive preferences allows us to do this.
- A matching μ is blocked by a group A of colleges and students if there exists another matching μ' such that for all students $s \in A$ and all colleges $c \in A$,
  - $\mu'(s) \in A$;
  - $\mu'(s) \succ_s \mu(s)$;
  - $s' \in \mu'(c)$ implies $s' \in A \cup \mu(c)$, i.e., every college A is matched at μ' to new students only from A, although it may continue to be matched with some of its "old" students from μ. (This differs from the definition of core.)
  - $\mu'(c) \succ_c \mu(c)$
- A matching is *group stable*, if it is not blocked by a group of any size.

Lemma 3. *When college preferences are responsive, a matching is group stable if and only if it is (pairwise) stable.*

Proof. Pairwise instability clearly implies group instability.

Now suppose μ is blocked via group A and outcome μ'. Then there must be a student $s \in A$ and a college $c \in A$ such that $s \in \mu'(c)$ but $s \notin \mu(c)$. So, s and c block μ (Otherwise it couldn't be that $\mu'(c) \succ_c \mu(c)$, since c has responsive preferences). $\square$

- Recall that
  - $\mathcal{S}$: set of (group) stable matchings.
  - $\mathcal{C}^W$: weak core.

– $\mathcal{C}^S$: (strong) core.

Lemma 4. *In the many-to-one matching model with strict preferences,*

$$\mathcal{S} = \mathcal{C}^S \subseteq \mathcal{C}^W.$$

College proposing deferred acceptance algorithm

Step 1:

- Each college c proposes to its q_c best choice students (if there are less acceptable choices then all of them).
- Each student rejects any unacceptable proposals and, if more than one acceptable proposal is received, "holds" the most preferred.
- If no proposals are rejected, and match each student to the college (if any) whose proposal she is "holding" and terminate the procedure.

Step k:

Any college who was rejected by l students at step $k-1$ makes $q_c - l$ new proposals to its most preferred acceptable students who has not yet rejected it. (If less acceptable choices remain, it proposes to all of its remaining acceptable students)

Each student "holds" her most preferred acceptable offer to date, and rejects the rest.

If no proposals are rejected, match each student to the college (if any) whose proposal she is "holding" and terminate the procedure.

Student proposing deferred acceptance algorithm

Step 1:

Each student s proposes to her best acceptable college.

Each college c rejects any unacceptable proposals and, if more than q_c acceptable proposals are received, "holds" the most preferred q_c group of students, if less than q_c acceptable proposals are received, it temporarily holds all of them.

If no proposals are rejected, and match each college to the student (if any) whose proposal it is "holding" and terminate the procedure.

Step k:

Any student who was rejected at step $k-1$ makes a new proposal to her most preferred acceptable college which has not yet rejected her.

Each college "holds" its most preferred q_c acceptable offers from this step and the ones it is holding from the previous step, and rejects the rest, if there are more than q_c of them. Otherwise, it "holds" all of the acceptable offers from this step and the ones from the previous step. It rejects all unacceptable offers from this step

If no proposals are rejected, match each student to the college (if any) whose proposal she is "holding" and terminate the procedure.

- Both of the above algorithms find stable matchings.
- Student proposing mechanism finds the student-optimal stable matching.

  College proposing mechanism finds the college-optimal stable matching.
- Lattice property of stable matchings still holds in a rather strong form.
- Student-optimal stable mechanism is still incentive compatible for students in dominant strategies. (But why?)

A related marriage market

- Replace college c by q_c positions of c denoted by $c^1, c^2, \ldots, c^{q_c}$. Each of these positions has c's preferences over individuals.

  Since each position c^i has a quota of 1, we do not need to consider preferences over groups of students.
- Each student's preference list is modified by replacing c, wherever it appears on his list, by the string $c^1, c^2, ..., c^{q_c}$, in that order.
- A matching μ of the many-to-one matching corresponds to a matching μ' in the related marriage market, in which the students in $\mu(c)$ are matched, in the order which

they occur in the preferences $\succ_c$, with the ordered positions of c that appear in the related marriage market. (If preferences are not strict, there will be more than one such matching.)

Lemma 5. *A matching of the many-to-one matching market is stable if and only if the corresponding matching of the related marriage market is stable.*

(Note: some results from the marriage model will translate immediately)

Theorem 11. *When all preferences over individuals are strict, the set of students enrolled and positions filled is the same at every stable matching.*

The proof is immediate via the similar result for the marriage market and the construction of the corresponding marriage market (Lemma 5).

So any college that fails to fill all of its positions at some stable matching will not be able to fill any more positions at any other stable matching. The next result shows that not only will such a college fill the same number of positions, but it will fill them with exactly the same students at any other stable matching.

Theorem 12 (Rural hospitals theorem)**.** *When preferences over individuals are strict, any college that does not fill its quota at some stable matching is assigned precisely the same set of students at every stable matching.*

Proof. This is proved by Lemma 6. □

Lemma 6. *Suppose colleges and students have strict individual preferences. Let μ and μ' be stable matchings for $(\mathbf{S}, \mathbf{C}, q, P)$, such that $\mu(c) \neq \mu'(c)$ for some $c \in \mathbf{C}$. Let $\hat{\mu}$ and $\hat{\mu}'$ be the stable matchings corresponding to μ and μ' in the related marriage market. If $\hat{\mu}(c^i) \succ_c \hat{\mu}'(c^i)$ for some position c^i of c, then $\hat{\mu}(c^i) \succsim_c \hat{\mu}'(c^i)$ for all positions c^i of c.*

Proof. It is enough to show that $\hat{\mu}(c^j) \succ_c \hat{\mu}'(c^j)$ for all $j > i$. So, suppose that this is false. Then, there exists an index j such that $\hat{\mu}(c^j) \succ_c \hat{\mu}'(c^j)$, but $\hat{\mu}'(c^{j+1}) \succsim_c \hat{\mu}(c^{j+1})$. Theorem 11 implies $\hat{\mu}'(c^j) \in \mathbf{S}$. Let $s' \equiv \hat{\mu}'(c^j)$. By the Decomposition Lemma, $c^j \equiv \hat{\mu}'(s') \succ_{s'} \hat{\mu}(s')$. Furthermore, $\hat{\mu}(s') \neq c^{j+1}$, since $s' \succ_c \hat{\mu}'(c^{j+1}) \succsim_c \hat{\mu}(c^{j+1})$ (where the first of these preferences follows from the fact that for any stable matching $\hat{\mu}'$ in the related marriage market, $\hat{\mu}(c^j) \succ_c \hat{\mu}'(c^{j+1})$ for all j). Therefore, c^{j+1} comes right after c^j in the preferences of s' (or any s) in the related marriage market. So, $\hat{\mu}$ is blocked via s' and c^{j+1}, contradicting (via Lemma 5) the stability of μ. □

Theorem 13. *Let preferences over individuals be strict, and let μ and μ' be stable matchings for $(\mathbf{S}, \mathbf{C}, q, P)$. If $\mu(c) \succ_c \mu'(c)$ for some college c, then $s \succ_c s'$ for all $s \in \mu(c)$ and $s' \in \mu'(c) - \mu(c)$. That is, c prefers every student in its assignment at μ to every student who is in its assignment at μ' but not at μ.*

Proof. Consider the related marriage market and the stable matchings $\hat{\mu}$ and $\hat{\mu}'$ corresponding to μ and μ'. Let $q_c = k$, so that the positions of c are $c^1, \ldots, c^k$. First, observe that c fills its quota under μ and μ' since, if not, Theorem 12 would imply that $\mu(c) = \mu'(c)$. So, $\mu'(c) - \mu(c)$ is nonempty subset of $\mathbf{S}$, since $\mu(c) \neq \mu'(c)$. Let $s' = \hat{\mu}'(c^j)$ for some position c^j such that s' is not in $\mu(c)$. Then, $\hat{\mu}(c^j) \neq \hat{\mu}'(c^j)$. By Lemma 6, $\hat{\mu}(c^j) \succ_c \hat{\mu}'(c^j) = s'$. The Decomposition Lemma implies $c^j \succ_{s'} \hat{\mu}(s')$. So the construction of the related marriage market implies $c \succ_{s'} \mu(s')$, since $\mu(s') \neq c$. Thus, $s \succ_c s'$ for all $s \in \mu(c)$ by the stability of μ, which completes the proof. □

- Consider a college c with $q_c = 2$ and preferences over individuals

$$s_1 \succ_c s_2 \succ_c s_3 \succ_c s_4.$$

Suppose that at various matchings 1-4, c is matched to

(matching 1) $\{s_1, s_4\}$

(matching 2) $\{s_2, s_3\}$

(matching 3) $\{s_1, s_3\}$

(matching 4) $\{s_2, s_4\}$

- Which matchings can be simultaneously stable for some responsive preferences over individuals?
- So long as all preferences over groups are responsive, matchings 1 and 2 cannot both be stable (Lemma 6), nor can matchings 3 and 4 (Theorem 13).

Theorem 14. *A stable matching procedure which yields the student-optimal stable matching makes it a dominant strategy for all students to state their true preferences.*

Proof. The result is immediate from the related marriage market. □

Theorem 15. *No stable matching mechanism exists that makes it a dominant strategy for all colleges to state their true preferences.*

Proof. Consider a college admission market $(\mathbf{C}, \mathbf{S}, q, P)$ with 3 colleges and 4 students. c_1 has a quota of 2 and both other colleges have a quota of 1. The preferences are

$$\begin{array}{llcll} c_1: & s_1 \succ s_2 \succ s_3 \succ s_4 & \qquad & s_1: & c_3 \succ c_1 \succ c_2 \\ c_2: & s_1 \succ s_2 \succ s_3 \succ s_4 & & s_2: & c_2 \succ c_1 \succ c_3 \\ c_3: & s_3 \succ s_1 \succ s_2 \succ s_4 & & s_3: & c_1 \succ c_3 \succ c_2 \\ & & & s_4: & c_1 \succ c_2 \succ c_3 \end{array}$$

The unique stable matching is

$$\mu = \begin{array}{ccc} c_1 & c_2 & c_3 \\ s_3 s_4 & s_2 & s_1 \end{array}.$$

But, if c_1 instead submitted the preferences

$$c_1: \quad s_1 \succ s_4,$$

the unique stable matching is

$$\mu' = \begin{array}{ccc} c_1 & c_2 & c_3 \\ s_1 s_4 & s_2 & s_3 \end{array}.$$

Note that $\mu' \succ_{c_1} \mu$. □

References

Roth, A. E. and Sotomayor, M. A. O. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press.