Introduction/ Basic Theory of Two-Sided Matching

Fuhito Kojima[1]

July 23, 2010
Mini Lectures at University of Tokyo

[1] Stanford University. http://sites.google.com/site/fuhitokojimaeconomics/.

What is Matching and Market Design?

Most traditional economics focuses on analyzing economy as it is.

Recently economists have been using economics to design institutions successfully, such as

1. student placement in schools, and
2. labor markets where workers.

The economics of "market design" analyzes and designs real-life institutions. A lot of emphasis is placed on concrete markets and details so that we can offer practical solutions.

Labor Markets: The case of American hospital-intern markets.

- Medical students in many countries work as residents (interns) at hospitals.
- In the U.S. more than 20,000 medical students and 4,000 hospitals are matched through a clearinghouse, called NRMP (National Resident Matching Program).
- Doctors and hospitals submit preference rankings to the clearinghouse, and the clearinghouse uses a specified rule to decide who works where.
- Some markets succeeded while others failed. What is a "good way" to match doctors and hospitals?

School Choice

- In many countries, children were automatically sent to a school in their neighborhoods.
- Recently, more and more cities in the United States and in other countries employ **school choice** programs: school authorities take into account preferences of children and their parents.
- Because school seats are limited (for popular schools), schools districts should decide who is admitted.
- How should school districts decide placements of students in schools?

Basic Setup

There are a set of doctors D and a set of hospitals H.

$\succ_d$: doctor d's strict preference over hospitals and being unmatched, $\emptyset$.
(we write $h \succeq_d h'$ if and only if $h \succ_d h'$ or $h = h'$.)

$\succ_h$: hospital h's strict preferences over subsets of doctors
(we write $D' \succeq_h D''$ iff $D' \succ_h D''$ or $D' = D''$.)

Assume each hospital's preference $\succ_h$ is **responsive with capacity** q_h (Roth 1985), i.e.,

1. the hospital's ranking over individual doctors is independent of her colleagues, and
2. any set of doctors exceeding q_h is unacceptable.

Matching

The outcome of the matching market is a **matching**, which specifies which doctor attends which hospital.

Formally, a matching μ is a mapping from D to $H \cup \{\emptyset\}$, with μ_d denoting a matching for doctor d.

For each hospital $h \in H$, let $\mu_h = \{d \in D | \mu_d = h\}$ be the set of doctors employed by h.

Stability

- Roughly speaking, a matching is stable if there are no individual players or pairs of players who can profitably deviate from (block) it. Formally,
- Matching μ is **individually rational** if $\mu_d \succeq_d \emptyset$ for every doctor d; for each hospital h, $|\mu_h| \leq q_h$ and $d \succ_h \emptyset$ for every $d \in \mu_h$.
- Matching μ is **blocked by a pair** d and h if each of them prefer each other to their partners under μ, that is, either
  1. $h \succ_d \mu_d$ and $d \succ_h d'$ for some $d' \in \mu_h$, or
  2. $h \succ_d \mu_d$ and $|\mu_h| \leq q_h$
- A matching is **stable** if it is individually rational and it is not blocked by any pair.
- (a note: the set of all stable matchings is equivalent to the core, and a stable matching is **Pareto efficient**.)

Stable matchings always exist

Theorem (Gale and Shapley 1962; RS Theorem 2.8)

There exists a stable matching in any one-to-one matching market.

- Gale and Shapley propose the (doctor-proposing) **deferred acceptance algorithm**:
- Given preferences of doctors and hospitals, conduct the following algorithm:

Step 1 : (a) Each doctor "applies" to her first choice hospital.
(b) Each hospital keeps the most preferred applicant (if s/he is acceptable) and rejects all other doctors.

Step $t \geq 2$: (a) Each doctor rejected in Step $(t-1)$ applies to her next highest choice.
(b) Each hospital considers both new applicants and the doctor (if any) held at Step (t-1), keeps the most preferred acceptable doctor from the combined set of doctors, and rejects all other doctors.

- Terminate when no more applications are made. Termination happens in finite time.

Example of DA algorithm

- Let $D = \{d_1, d_2, d_3\}, H = \{h_1, h_2\}$, and their preferences be

$$\begin{aligned}
\succ_{d_1} &: h_1, h_2, \\
\succ_{d_2} &: h_1, \\
\succ_{d_3} &: h_2, h_1, \\
\succ_{h_1} &: d_3, d_2, d_1, \\
\succ_{h_2} &: d_1, d_3.
\end{aligned}$$

- Follow steps of the DA algorithm (I recommend you do it with a piece of paper).
- The resulting matching $\mu = \{(d_1, h_2), (d_2, \emptyset), (d_3, h_1)\}$ is stable (verify it!).

Proof of Theorem (A stable matching always exists)

The proof is very simple.

1. The resulting matching μ of DA is individually rational because at each step of the algorithm, no doctor applies to an unacceptable hospital and no hospital holds only acceptable doctors and up to its quota.
2. μ is not blocked by any pair because: Suppose $h \succ_d \mu_d$ for some d and h. This means that d applied to h and was rejected by h at some step of DA. Since d's tentative match only improves as the algorithm proceeds, at the match μ_h at the end of DA, all positions of h are filled with doctors more preferred by h to d. So h is not interested in blocking μ with d.

Mechanisms in real markets

1. Stability is theoretically appealing, but does it matter in real life?
2. Roth (1984) showed that the NIMP algorithm is equivalent to a (hospital-proposing) DA algorithm, so NIMP produces a stable matching.
3. Roth (1991) studied British medical match, where different regions use different matching mechanisms. He found that stable mechanisms are successfully used (and is still in use) but most unstable mechanisms were abandoned after a short period of time.
4. In school choice, stability means "no justified envy": no student is placed in a less preferred school to another school where a student with lower priority is assigned. NYC and Boston recently adopted DA in order to, among other things, to eliminate such unfair assignment.

Mechanisms in real markets

Market	Stable	Still in use
NRMP	yes	yes (new design 98-)
Edinburgh ('69)	yes	yes
Cardiff	yes	yes
Birmingham	no	no
Edinburgh ('67)	no	no
Newcastle	no	no
Sheffield	no	no
Cambridge	no	yes
London Hospital	no	yes
Medical Specialties	yes	yes (1/30 no)
Canadian Lawyers	yes	yes
Dental Residencies	yes	yes (2/7 no)
Reform rabbis	yes	yes
NYC highschool	yes	yes

Improving Efficiency in Matching Markets with Regional Caps: The Case of The Japan Residency Matching Program

Yuichiro Kamada[2] Fuhito Kojima[3]

July 23, 2010

[1]Stanford University. http://sites.google.com/site/fuhitokojimaeconomics/.
[2]Harvard University.
[3]Stanford University.

Overview

Geographical distribution of medical doctors is a contentious issue in health care.

Many hospitals in rural areas do not attract enough medical residents to meet their demands.

Previous literature on stable matching suggests that a solution is elusive: the rural hospital theorem (Roth 1986).

Geographical Distribution of Residents in Japan

Japanese residency matching started in 2004 as part of a comprehensive reform of the medical residency program.

1. Prior to the reform, departments in university hospitals (called "ikyoku") had de facto power to allocate doctors.
2. The new system introduced a matching mechanism using the deferred acceptance algorithm by Gale and Shapley (1962).

Critics say that many rural hospitals fill fewer positions in the new matching mechanism.

Japanese government modified the system (JRMP mechanism):

1. The government sets a "regional cap" on each prefecture.
2. Reduce the capacity of each hospital so that the sum of the reduced capacities of hospitals of the region to equal the regional cap.
3. Then implement the deferred acceptance algorithm.

Main Results

This project

1. shows that the JRMP mechanism may result in avoidable inefficiency and instability,
2. points out the standard stability concept may be inadequate and formalizes several stability concepts under regional caps,
3. proposes the **flexible deferred acceptance mechanism** which
    1. improves efficiency and generates stable matchings while meeting the regional caps,
    2. is (group) strategy-proof for doctors.

$\rightarrow$ A potentially superior mechanism!

Basic Setup

There are a set of doctors D and a set of hospitals H.

$\succ_d$: doctor d's strict preference over hospitals and being unmatched, $\emptyset$.
(we write $h \succeq_d h'$ if and only if $h \succ_d h'$ or $h = h'$.)

$\succ_h$: hospital h's strict preferences over subsets of doctors
(we write $D' \succeq_h D''$ iff $D' \succ_h D''$ or $D' = D''$.)

Assume each hospital's preference $\succ_h$ is **responsive with capacity** q_h (Roth 1985), i.e.,

1. the hospital's ranking over individual doctors is independent of her colleagues, and
2. any set of doctors exceeding q_h is unacceptable.

A **matching** μ is a mapping from D to $H \cup \{\emptyset\}$, with μ_d denoting a matching for doctor d.

For each hospital $h \in H$, let $\mu_h = \{d \in D | \mu_d = h\}$ be the set of doctors employed by h.

Model of Regions

Each hospital belongs to exactly one **region** $r \in R$.

For each region r, there is a **regional cap** q_r (a positive integer).

A matching is **feasible** if $|\mu_{H_r}| \leq q_r$ for all $r \in R$, where H_r is the set of hospitals in region r and $\mu_{H_r} = \cup_{h \in H_r} \mu_h$.

This requirement distinguishes the environment from the standard model without regional caps.

The Deferred Acceptance (DA) Algorithm

Gale and Shapley (1962) consider a model with no binding regional cap, i.e., $q_r > |D|$ for every $r \in R$, and propose the **(doctor-proposing) deferred acceptance algorithm**. Start from a matching in which no one is matched.

Application Step:
Choose a doctor who is currently unmatched, and let her apply to her most preferred hospital that has not rejected her so far (if any).

Acceptance/Rejection Step:
Each hospital considers the combined pool of the tentatively matched doctors and the new applicant (if any). Specifically, the hospital chooses its most preferred acceptable doctors up to its capacity (if they exist) and rejects everyone else.

The algorithm terminates at a step in which no rejection occurs, producing a matching.

Why Use DA?

The DA mechanism has good properties:

1. DA produces a **stable matching**: there is no mutually profitable deviation by a doctor and a hospital.
    1. Stability $\iff$ Core.
    2. Empirical and experimental evidence that stability is important for the persistence of matching mechanisms (Roth 1984, 1991, Kagel and Roth 2000).
2. DA produces an efficient matching (because it is stable).
3. DA is (group) **strategy-proof** for doctors (Dubins and Freedman 1981, Roth 1982): reporting true preferences is a dominant strategy for every doctor.
4. DA is not strategy-proof for hospitals, but incentives for manipulation become small in large markets (Roth and Peranson 1999, Kojima and Pathak 2009).

The JRMP Mechanism

In Japan, government imposes a **target capacity** $\bar{q}_h \leq q_h$ for each hospital h such that $\sum_{h \in H_r} \bar{q}_h \leq q_r$ for each region $r \in R$.

The **JRMP mechanism** implements the deferred acceptance mechanism, except that it uses the target capacity instead of the hospital's actual capacity as input.

Idea: In order to satisfy regional caps, simply force hospitals to be matched to a smaller number of doctors than their real capacities, but otherwise use the standard deferred acceptance algorithm.

Note: In (most of) today's talk, the target capacities are exogenously given. In the current Japanese system, the target is decided by reducing the capacity of each hospital proportionately to equalize the total capacity to the regional cap.

But does the JRMP mechanism inherit good properties of DA?

JRMP May Produce an Inefficient Matching

There are two hospitals h_1, h_2 in one region with regional cap 10.

Each hospital has a capacity of 10 and a target capacity of 5.

There are 10 doctors, $d_1, \ldots, d_{10}$ such that

$$\begin{aligned} & d_1 \succ_h d_2, \succ_h \ldots \succ_h d_{10} \succ_h \emptyset, \text{ for both hospitals,} \\ & d_1, d_2, d_3 \text{ find only } h_1 \text{ acceptable,} \\ & d_4, \ldots, d_{10} \text{ find only } h_2 \text{ acceptable.} \end{aligned}$$

The JRMP mechanism produces

$$\begin{aligned} \mu_{h_1} &= \{d_1, d_2, d_3\} \\ \mu_{h_2} &= \{d_4, d_5, d_6, d_7, d_8\}. \end{aligned}$$

This matching is inefficient.

Stability

Clearly, the JRMP matching may be unstable in the standard sense.

We introduce a new stability concept that generalizes the standard stability notion to the case with regional caps.

Definition

A matching μ is **weakly stable** if it is feasible and

1. **Individual rationality**: $\mu_d \succeq_d \emptyset$ for each $d \in D$; $d \succeq_h \emptyset$ for all $h \in H$ and $d \in \mu_h$.
2. **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
   1. $\emptyset \succ_h d$.
   2. $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
   3. $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$ and $d' \succ_h d$ for all $d' \in \mu_h$.

The last part is the only difference from the standard notion.

Only "weak" stability, because movement to a hospital with a vacancy from a hospital in the same region is precluded.

JRMP May Produce an Unstable Matching

The same example as before: There are two hospitals h_1, h_2 in one region with regional cap 10.

Each hospital has a capacity of 10 and a target capacity of 5.

There are 10 doctors, $d_1, \ldots, d_{10}$ such that

$$d_1 \succ_h d_2, \succ_h \ldots \succ_h d_{10} \succ_h \emptyset, \text{ for both hospitals,}$$
$$d_1, d_2, d_3 \text{ find only } h_1 \text{ acceptable,}$$
$$d_4, \ldots, d_{10} \text{ find only } h_2 \text{ acceptable.}$$

The JRMP mechanism produces

$$\mu_{h_1} = \{d_1, d_2, d_3\}$$
$$\mu_{h_2} = \{d_4, d_5, d_6, d_7, d_8\}.$$

This matching is not stable.

The Flexible DA Mechanism

We define the **flexible deferred acceptance algorithm** below, given target capacity profile $(\bar{q}_h)_{h \in H}$. Start from a matching in which no one is matched.

Application Step:
Choose a doctor who is currently unmatched, and let her apply to her most preferred hospital that has not rejected her so far (if any).

Acceptance/Rejection Step:
For each region, each hospital in the region chooses from the combined applicant pool of the tentatively matched doctors and the new applicant (if any): For each region,

1. First, each hospital in the region chooses its most preferred acceptable applicants up to its target (if they exist).
2. Then, one by one, each hospital in the region takes turns (following a fixed order) to choose the most preferred remaining applicant until (i) the regional quota is filled or (ii) the capacity of the hospital is filled or (iii) no doctor remains to be matched.

Example of flexible DA

The same example as before: There are two hospitals h_1, h_2 in one region with regional cap 10.

Each hospital has a capacity of 10 and a target capacity of 5.

There are 10 doctors, $d_1, \ldots, d_{10}$ such that

$$d_1 \succ_h d_2, \succ_h \ldots \succ_h d_{10} \succ_h \emptyset, \text{ for both hospitals,}$$
$$d_1, d_2, d_3 \text{ find only } h_1 \text{ acceptable,}$$
$$d_4, \ldots, d_{10} \text{ find only } h_2 \text{ acceptable.}$$

The flexible DA mechanism produces

$$\mu_{h_1} = \{d_1, d_2, d_3\}$$
$$\mu_{h_2} = \{d_4, d_5, d_6, d_7, d_8, d_9, d_{10}\},$$

which is weakly stable and efficient.

Stability

Theorem

The flexible deferred acceptance mechanism produces a stable matching for any input.

Intuition:

Unlike JRMP, the target capacity of each hospital is not rigid.

As long as the regional cap is not violated, hospitals can tentatively accept doctors beyond the target capacities.

Like the DA, an acceptable doctor rejected from a more preferred hospital was rejected either because there are enough better doctors in that hospital, or regional quota was filled by other doctors.
$\rightarrow$ The doctor cannot form a blocking pair!

Efficiency

Theorem

Any stable matching is efficient.

Note: This result is well-known when there is no regional cap, and is a straightforward implication of the fact that stability is equivalent to core.

But with regional caps, there is no obvious way to define the core. Fortunately the statement still goes through.

Corollary

The flexible deferred acceptance mechanism produces an efficient matching for any input.

Strong Stability

Our weak stability concept may be problematic because blocking within a region does not violate regional caps.

The next definition formalizes a more strict concept of stability taking this issue into consideration.

Definition

A matching μ is **strongly stable** if it is feasible and

1. **Individual rationality**,
2. **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
   1. $\emptyset \succ_h d$.
   2. $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
   3. $\mu_d \notin H_r$ and $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.

Strongly Stable Matchings May Not Exist

There is one region with regional cap of one, with two hospitals h_1 and h_2 with capacity one each and two doctors, d_1 and d_2, with preferences

$$\succ_{h_1}: d_1, d_2, \qquad \succ_{h_2}: d_2, d_1,$$

$$\succ_{d_1}: h_2, h_1, \qquad \succ_{d_2}: h_1, h_2.$$

1. No matching in which two doctors are matched is feasible because it violates the regional cap.
2. If no doctor is matched, then there is a blocking pair (d_1 and h_1 for example).
3. A matching where $\mu_{h_1} = \{d_1\}$. $\rightarrow (d_1, h_2)$ is a blocking pair.
4. A matching where $\mu_{h_1} = \{d_2\}$. $\rightarrow (d_1, h_1)$ is a blocking pair (h_1 can reject d_2 to be paired with d_1).
5. $\mu_{h_2} = \{d_2\}$ and $\mu_{h_2} = \{d_1\}$ is not strongly stable (symmetric argument).

Stability

Definition

A matching μ is **stable** with respect to a target capacity $(\bar{q}_h)_{h\in H}$ if it is feasible and

1. **Individual rationality**,
2. **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
   1. $\emptyset \succ_h d$.
   2. $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
   3. $\mu_d \notin H_r$ and $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$, $\leftarrow$ same as strong stability
   4. $\mu_d \in H_r$, $|\mu_h| \geq \bar{q}_h$, $|\mu_h| + 1 - \bar{q}_h > |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}$, $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$. $\leftarrow$ new!

Theorem

The flexible deferred acceptance mechanism produces a stable matching for any input.

Incentives

Theorem

The flexible DA mechanism is (group) strategy-proof for doctors: Truthtelling is a dominant strategy for every doctor.

A (very rough) intuition: a doctor doesn't need to give up trying for her first choice because, even if she is rejected, she will be able to apply to her second choice etc. The deferred acceptance guarantees that she will be treated equally if she applies to a position later than others.

Truthtelling is not necessarily a dominant strategy for hospitals.

Impossibility theorem (Roth 1982): There is no strategy-proof and stable mechanism.

Failure of Rural Hospital Theorem

The rural hospital theorem fails: The set of unmatched doctors and hospitals can differ across stable matchings.

There are two regions r and r' with regional cap of one each. Hospitals h_1 and h_2 are in r and h_3 is in r' with capacity one each.

Preferences are

$$\succ_{h_1}: d_1, d_2, \qquad \succ_{h_2}: d_2, d_1, \qquad \succ_{h_3}: d_2,$$

$$\succ_{d_1}: h_1, h_2, \qquad \succ_{d_2}: h_2, h_3.$$

One stable matching matches d_1 to h_1 and d_2 to h_3. Another stable matching matches d_2 to h_2 only.

Given this, design of the mechanism may influence geographical distributions of doctors.

Our Results Put in Context

Our direct contribution is practical: proposing a better mechanism for the Japanese residency matching market.

Potential applications include:

1. Residency markets in other countries.
2. U.S. medical resident markets: ACGME decides total numbers of residents in each specialty.
3. Student placement in public schools: multiple school programs share one school building.

Theoretically, we propose a new model of matching with regional caps. New stability concepts are defined and analyzed.

Methodologically, this project tries to advance market design (economic engineering) to solve practical design problems.

Conclusion

This paper shows that the JRMP can be improved upon by another mechanism, flexible DA.

In so doing, we proposed a new matching problem with regional caps and introduced appropriate stability concepts.
$\longrightarrow$ the model may be useful more generally.

Other Issues and Future Research:

1. Alternative policy goals $\longrightarrow$ generalized flexible DA algorithms.
2. Empirical study or simulation.
3. Other applications?