The problem sets and exams build on the material from the notes. You can rely on the notes during exams.

This note introduces the economic model of two-sided matching, the concept of stable matchings, and the deferred acceptance algorithm (DA) that finds stable matchings. We use the algorithm to show that a stable matching always exist.

1 The Model of Two-sided Matching

The two sided matching model was introduced by Gale and Shapley 1962 paper in *American Mathematical Monthly*. Their exposition is truly outstanding, and I recommend reading their original paper to all who are interested (their paper is posted under Week 1 on our website).

Participants There are two types of agents in the model, following the tradition let us call them men and women. For simplicity let us assume...

*I would like to thank Jonathan Levine and Mallesh Pai for sharing with me their notes on the topic.*
that there are as many men as women. There is thus a set of \( n \) men \( M = \{m_1, m_2, \ldots, m_n\} \), and a set of \( n \) women \( W = \{w_1, w_2, \ldots, w_n\} \).

**Preferences** Each woman has a strict preference over all men in \( M \) plus the option to stay single \( \emptyset \). For instance, woman \( w_1 \) may prefer to be with man \( m_1 \) over \( m_2 \) over \( m_3 \), but may prefer being unmarried to any of the other men. We will denote such a preference as \( m_1 \succ m_2 \succ m_3 \succ \emptyset \ldots \).

Similarly, each man has preferences over the set of women \( W \) and staying single \( \emptyset \).

**Matchings** A matching is a list of pairs (couples) \((m_j, w_k)\) and unmatched (single) agents such that each man and each woman appears exactly once in this list.

## 2 Stability

A matching is **stable** if there is no:

1. **Blocking Individual** There is no individual matched to an unacceptable partner. In other words, there is no one who would rather be single than remain married to their current partner.

2. **Blocking Pair** There is no pair (a man and a woman) who would rather marry each other than stay with their current assignments.

Why consider stability? Recall the residency matching, NRMP. If the outcome of a matching process is unstable, then people may want to look to contract outside the system, or re-contract after the system. Roth JPE (1984) (see website) discusses empirical evidence that matching systems that generate stable matchings last while system that generate unstable matchings do not.
Example 1. Consider a simple setting with 2 men and 2 women. The preferences are as shown below:

<table>
<thead>
<tr>
<th></th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$w_1$</th>
<th>$w_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st choice</td>
<td>$w_1$</td>
<td>$w_1$</td>
<td>$m_2$</td>
<td>$m_1$</td>
</tr>
<tr>
<td>2nd choice</td>
<td>$w_2$</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td>$m_2$</td>
</tr>
</tbody>
</table>

The matching $\{(m_1, w_2), (m_2, w_1)\}$ is unstable because $m_1$ is a blocking individual. The matching $\{m_1, w_1, (m_2, w_2)\}$ is unstable because $m_1$ and $w_1$ are a blocking pair; $m_2$ and $w_1$ are another blocking pair.

The matching $\{m_1, w_2, (m_2, w_1)\}$ is stable. Indeed, there is no blocking individual as $m_1$ and $w_2$ are already single, and each of $m_2$ and $w_1$ prefer being together to being single. Furthermore, there are no blocking pair as $m_2$ and $w_1$ are each other first choices and would not participate in any blocking pair; this leaves $m_1$ and $w_2$ to form a potential blocking pair but they are not a blocking pair either as $w_2$ is not acceptable to $m_1$.

Exercise 1. Suppose preferences instead are:

<table>
<thead>
<tr>
<th></th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$w_1$</th>
<th>$w_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w_1$</td>
<td>$w_1$</td>
<td>$m_1$</td>
<td>$m_1$</td>
</tr>
<tr>
<td></td>
<td>$w_2$</td>
<td>$m_2$</td>
<td>$m_2$</td>
<td>$m_2$</td>
</tr>
</tbody>
</table>

The matching $(m_1, w_2)$ and $(m_2, w_1)$ is unstable because $(m_1, w_1)$ form a blocking pair. Is there a stable matching? If yes, what matching is stable?

Here is a similar exercise:

Exercise 2. Suppose preferences instead are:

<table>
<thead>
<tr>
<th></th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$w_1$</th>
<th>$w_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w_1$</td>
<td>$w_1$</td>
<td>$m_2$</td>
<td>$m_1$</td>
</tr>
<tr>
<td></td>
<td>$w_2$</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td>$m_2$</td>
</tr>
</tbody>
</table>

Check that the matching $(m_1, w_2)$ and $(m_2, w_1)$ is stable. One way to do it
is to check for each individual whether he is blocking or not, and for each pair of a man and a woman who are not matched to each other whether they form a blocking pair.

3 The Deferred Acceptance Algorithm (DA)

Gale and Shapley (1962) proposed a simple algorithm to find a stable matching. The algorithm is often referred to as deferred acceptance algorithm (DA). The algorithm proceeds in rounds. Each round consists of two steps:

1. **Proposing:** Each man proposes to his most preferred acceptable woman that did not reject him in a prior round.

2. **Rejections:** Each woman holds onto her most preferred acceptable man (if any), and rejects all other proposals.

The algorithm ends when no rejections are made in a round. At this point each man whose proposal is held by a women is matched with the woman. All remaining men and women remain unmatched.

Question: does the algorithm always terminate? Yes, a man who was rejected by a woman is not going to apply to her again; thus there is only a finite number of rejections that can be made.

The standard way to run the algorithm is to first solicit preferences from all agents, and then replicate the sequence of proposals and acceptances/rejections. It is also possible to run the algorithm so that the agents themselves make the proposals and accept/reject them.

**Exercise 3.** Consider an economy with 4 men and 4 women. Pick preferences for each agent such that they all prefer being married to being single. Run the deferred acceptance algorithm to end up with a matching. Is the matching you find stable?
4 Deferred Acceptance and Stable Matchings

**Theorem 1.** *The deferred acceptance algorithm always terminates and the resulting matching is stable.*

*Note:* on the exams we will work with examples; you will not be required to prove results on the exams. The proof below—and the proofs in general—is thus optional. If you are interested in some harder questions related to this material, you may want to try to work out the proof of the theorem before reading the proof proposed below.

*Proof.* We have already seen that DA terminates. To verify that the resulting matching is stable we need to verify that there are no blocking individuals and there are no blocking pairs.

*Blocking Individuals:* No man could end up with a woman who is unacceptable to him, since he wouldn’t have proposed to her in the first place. No woman could end up with an unacceptable man because at no stage would she have tentatively accepted his proposal.

*Blocking Pairs:* Suppose by way of contradiction that there is a man $m$ and a woman $w$ who would have preferred to be with each other than their present partners. This implies that $m$ proposed to $w$ in some round, and was rejected by her. At the round show rejected him, woman $w$ must have had a proposal from a man she prefers over $m$. In subsequent rounds of DA she would keep this proposal or rejected it in favor of even better proposals. Thus, at the end of the algorithm woman $w$ is matched to a man she prefers over $m$. Hence, we have a contradiction as $w$ is not willing to form a blocking pair with $m$ after all.

As an immediate corollary, we conclude that

**Corollary 1.** *For any profile of preferences of men $M$ and women $W$, there exists a stable matching.*
5 There Might Be More than One Stable Matching

There are preference profiles that allow multiple stable matches. Consider for instance the following preference profile.

\[
\begin{array}{c|c|c|c}
  m_1 & m_2 & w_1 & w_2 \\
  w_1 & w_2 & m_2 & m_1 \\
  w_2 & w_1 & m_1 & m_2 \\
\end{array}
\]

This preference profile leads to two stable matchings: one can be obtained by running the standard deferred acceptance algorithm, with men proposing. We sometimes refer to this version of DA as *Men-Proposing Deferred Acceptance* algorithm. To obtain the other we can run a symmetric twin of the men-proposing deferred acceptance, and have women propose in step 1 of each round, and men hold onto some proposals and reject others in step 2. We will refer to this twin version of DA as *Women-Proposing Deferred Acceptance* algorithm.

In this particular example, the stable matching generated by the women-proposing DA is preferred by all the women to the stable matching generated by the men-proposing DA. Similarly, the stable matching generated by the men-proposing DA is preferred by all the men to the stable matching generated by the women-proposing DA.

**Exercise 4.** [Optional and hard] Is the above observation generally true? Of course, it can happen that some women get same partners under both men-proposing and women-proposing DA. For women that get different partners under the two procedures, do all of them prefer the outcome of the women-proposing DA?