CS386 Scribe Notes: Stable Marriage

Shahzaib Awan

November 10, 2026

Overview: Today's lecture covered the Stable Marriage problem, including the Deferred Acceptance algorithm and proofs of stability and optimality. I have added simple English explanations alongside the formal math to make these notes easier to follow.

1 Introduction to Stable Marriage

The Setup: We are given a set of men M and a set of women W. Every person has a ranked list of some subset of the people in the opposite sex. Anyone not ranked is deemed unacceptable.

Goal: Compute a stable matching.

A matching is considered **stable** if two conditions hold:

- No one is matched to someone they deem unacceptable.
- No man and woman prefer each other over their current assignment. (In simple terms, no two people would want to break their current matches to run away together).

This raises a few important questions:

1. Is it even guaranteed that there is a stable matching?
2. Even if there is one, how do we find it?
3. If there are multiple stable assignments, how do I pick which one to use?

Idea 1: Start with any matching, "fix" the instabilities

What if we start with an arbitrary matching and just swap partners as we find instabilities? As shown on the board, this approach does not always work because fixing one instability might create another, leading to an infinite cycle.

Consider the following preference rankings for 3 men and 3 women:

M_1	M_2	M_3	W_1	W_2	W_3
W_2	W_3	W_1	M_2	M_3	M_1
W_3	W_1	W_2	M_3	M_1	M_2
W_1	W_2	W_3	M_1	M_2	M_3

If we try to fix instabilities (underlined below) one by one, we get trapped in a loop where the matchings repeat forever:

$$\begin{aligned} S_1 :&\{(\underline{M_1}, W_1), (M_2, W_2), (M_3, \underline{W_3})\} \\ \hookrightarrow S_2 :&\{(\underline{M_1}, W_3), (M_2, \underline{W_2}), (M_3, W_1)\} \\ \hookrightarrow S_3 :&\{(M_1, \underline{W_2}), (M_2, W_3), (\underline{M_3}, W_1)\} \\ \hookrightarrow S_4 :&\{(M_1, W_1), (\underline{M_2}, W_3), (M_3, \underline{W_2})\} \\ \hookrightarrow S_1 &\text{ (The cycle repeats)} \end{aligned}$$

Because of this cyclic behavior, we need a more structured strategy.

2 The Deferred Acceptance Algorithm

To guarantee a stable matching, we use the Deferred Acceptance Algorithm (often called the Gale-Shapley algorithm).

The Rules:

- **Men:** Sequentially, men make proposals to the woman they have not yet proposed to whom they rank the highest.
- **Women:** Women react to the proposals based on the following rules:
  1. Immediately reject a man she doesn't find acceptable.
  2. If she's single and finds him acceptable, tentatively accept.
  3. If a more preferable man proposes, reject the current partner and tentatively accept the new one.

Algorithm Evolution and Stability

To see how this evolves and reaches stability, let's trace the algorithm using the preferences from the earlier table where the men are proposing:

- **Round 1 (Men Proposing):**
  - M_1 proposes to his top choice, W_2. W_2 tentatively accepts him (he is on her list).
  - M_2 proposes to his top choice, W_3. W_3 tentatively accepts him.
  - M_3 proposes to his top choice, W_1. W_1 tentatively accepts him.
- **Result:** Because no two men proposed to the same woman, there are no rejections. The algorithm terminates immediately. If, for example, M_1 and M_2 had both proposed to W_3, she would have kept her preferred man and rejected the other, forcing the rejected man to propose to his next best choice in Round 2. The matching evolves in this way until no one is rejected.

Depending on who proposes (men vs. women), the algorithm can produce different matchings. On the board, the two resulting matchings were:

- **Proposing Men (μ_M):** $\{(M_1, W_2), (M_2, W_3), (M_3, W_1)\}$
- **Proposing Women (μ_W):** $\{(M_1, W_3), (M_2, W_1), (M_3, W_2)\}$

Proof of Stability

Why does this algorithm work? Let's prove it by contradiction.

- Suppose a man m prefers a woman w to his actual assigned matching $\mu_M(m)$.
- Mathematically, we write this as $w >_m \mu_M(m)$.
- Because men propose in decreasing order of preference, m must have proposed to w before proposing to $\mu_M(m)$.
- At some point, w must have rejected m for some other man m'. This means $m' >_w m$.
- Since women's matches only get better over time, her final match $\mu(w)$ must be at least as good as m'.
- Therefore, $\mu(w) \geq_w m' >_w m$.
- Because w strongly prefers her final match over m, she will not run away with him. Thus, the system is strictly stable.

3 Optimality of the Matchings

A fascinating property of Deferred Acceptance is that the proposing side gets the best possible outcome, while the receiving side gets the worst.

- In μ_M, every man ends up with his **best possible valid partner**. That is, for all possible stable matchings μ, $\mu_M(m) \geq \mu(m)$.
- Every single woman ends up with her **worst possible valid partner**.

Proof Sketch for Male Optimality

Let's prove that no man is ever rejected by his best valid stable partner.

- Let m be the **first man** in the entire algorithm to be rejected by a woman w who is a valid partner for him (meaning there exists some stable matching μ where $\mu(m) = w$).
- Let m' be the man that w chose instead of m (so she rejected m for m').
- Because m is the *first* man rejected by a valid partner, m' has not yet been rejected by any of his valid partners.
- If m' prefers w over his own match in μ ($w >_{m'} \mu(m')$), then μ would not be stable because m' and w would run away together.
- Therefore, it must be that $\mu(m') \geq_{m'} w$.
- But if m' prefers $\mu(m')$ over w, he would have proposed to $\mu(m')$ first and been rejected by her *before* proposing to w. This contradicts our assumption that m was the first man rejected by a valid partner!

Comparing Stable Matchings

If μ and ν are stable matchings, then $\mu >_M \nu \iff \nu >_W \mu$. (Note: We say $\mu \geq \nu$ if for all people p in the set, $\mu(p) \geq_p \nu(p)$ and there exists at least one person p such that $\mu(p) >_p \nu(p)$).

Proof:

- Suppose $\mu >_M \nu$ (the men prefer μ) and suppose there exists a woman w who also prefers μ, so $\mu >_w \nu$.
- This means her match in both systems is different: $\mu(w) \neq \nu(w)$.
- Let $m = \mu(w)$. Since her match is different, m's match is also different: $\mu(m) \neq \nu(m)$.
- By our first assumption ($\mu >_M \nu$), m prefers μ over ν, meaning $w >_m \nu(m)$.
- By our second assumption ($\mu >_w \nu$), w prefers μ over ν, meaning $m >_w \nu(w)$.
- Because both m and w prefer each other over their assigned partners in ν, the matching ν cannot be stable. This contradiction proves the inverse relationship: what is better for the men is strictly worse for the women.

4 Changes in Preferences (Strategy-Proofness)

Finally, we looked at what happens when preference rankings change slightly, which addresses whether participants can "game" the system by lying about their preferences.

Let $<$ be the honest preference lists, and let $<'$ be the manipulated preference lists. These lists are identical except for the preferences of a specific coalition of people C ($<_p$ for all $p \in C$).

Then, for any stable matchings μ generated using the honest preferences $<$, there exists a stable matching μ' generated using the manipulated preferences $<'$, such that there is at least one person $p \in C$ where:

$$\mu \geq_p \mu'$$

In Simple English: You cannot successfully manipulate the Deferred Acceptance algorithm to strictly benefit everyone in a group. If a coalition tries to lie about their preference lists to get better matches, the math guarantees that at least one person in that coalition will end up with an outcome that is **equal to or worse** than what they would have gotten by just being honest.