# DS 102 Discussion 9

Wednesday, November 4, 2020

In Lecture 20 this week, we introduce the **stable matching problem** and a classic algorithm for solving it, the **Gale-Shapley** algorithm. In this discussion, we'll look at a **truthfulness** implication of the Gale-Shapley algorithm. That is, can proposers or proposees obtain a better outcome via the Gale-Shapley algorithm if they lie about their true preferences? Here, we'll show that no proposer has an incentive to lie about their preferences. Whether or not a strategy like Gale-Shapley is truthful helps us understand the consequences that might arise when it is deployed in real-world matching markets.

First, we briefly review relevant concepts covered in Lecture 20. We consider a group of proposers and a group of proposees, where each proposer has a strict preference order over all the proposees and each proposee has a strict preference order over all the proposers. A **matching** $\mu$ is a one-to-one mapping from a subset of the proposers to a subset of the proposees, where $\mu(i)$ refers to the individual matched to individual $i$. A matching $\mu$ is **unstable** if there exists a proposer and a proposee who are not matched to each other in $\mu$, but prefer each other to their partners in $\mu$. We assume every individual prefers being matched to being unmatched. Otherwise, the matching is called **stable**.

The Gale-Shapley algorithm for finding a stable matching proceeds with the following steps.

1. Initially, all proposers and proposees are unmatched.

2. Each proposer proposes to their most preferred proposee who has not rejected them yet. If a proposer has been rejected by all proposees, they give up and cease proposing.

3. Each proposee is tentatively matched to their favorite among their proposers for this round, and rejects the rest.

4. Repeat Steps 2 and 3 until there is a round in which there are no rejections. The tentative matches then become final.

1. **Proposers have no incentive to lie.** In this problem, we'll do a proof sketch of the following claim.

   *Claim.* Let $\mu$ denote the stable matching produced by the Gale-Shapley algorithm. Suppose that a proposer, $m_0$, lies about their preferences for the proposees (*i.e.*, reports a preference order other than their true one). Then there is no stable matching for the modified preference profile where $m_0$ obtains a better match than in $\mu$ (according to their true preference order).

   (a) Using the following Lemma, prove the claim above with a proof by contradiction.

   *Lemma.* Let $\mu$ be the stable matching produced by the Gale-Shapley algorithm and let $\nu$ be another matching. Denote by S the set of proposers who prefer their match in $\nu$ to their match in $\mu$:

$$S = \{m \mid m \text{ prefers } \nu(m) \text{ to } \mu(m)\}.$$

If $S$ is non-empty, then $\nu$ is unstable due to a pair $(m, w)$ where $m \notin S$.

**Solution:** Let $\nu$ be any matching where $m_0$ has a better match than in $\mu$, according to their true preference order. The pair $(m, w)$ identified in the Lemma makes $\nu$ unstable for the modified preference profile. To see this, note that only $m_0$'s preference order is different in the modified preference profile. Also note that $m \neq m_0$, since $m_0 \in S$ and $m \notin S$, and that $m$ and $w$'s preferences orders have not changed. Therefore, since $(m, w)$ is unstable for the original preference order, it is also unstable for the modified preference order. Since $\nu$ is unstable under the modified preference order, the Gale-Shapley algorithm will not obtain it, so $m_0$ cannot benefit from lying about their preferences.

(b) To complete our proof of the claim, we just need to prove the Lemma. This can be done by considering two different cases. For simplicity, here we'll only be concerned about proving one case (don't worry about how to show the other case).

We consider the case when $\mu(S) \neq \nu(S)$. That is, the set of proposees that the proposers in $S$ are matched to are different in $\mu$ and $\nu$ (see Figure 1). Let $w \in \nu(S) \setminus \mu(S)$ and let $m = \mu(w)$. Show that $(m, w)$ makes $\nu$ unstable.

**Solution:** First, note that $m \notin S$, since $w \notin \mu(S)$. Therefore, by definition of $S$, $m$ prefers $w = \mu(m)$ to $\nu(m)$. Second, in the execution of the Gale-Shapley algorithm to obtain $\mu$, since $\nu(w) \in S$ we know that $\nu(w)$ proposed to $w$ before they proposed to their final partner in $\mu$. However, $\nu(w)$ does not end up partnered with $w$, which means $w$ must have rejected $\nu(w)$. That means that $w$ must prefer $m = \mu(w)$, their final partner in $\mu$, over $\nu(w)$. That is, $w$ and $m$ both prefer each other to their partners in $\nu$, so $(m, w)$ makes $\nu$ unstable.

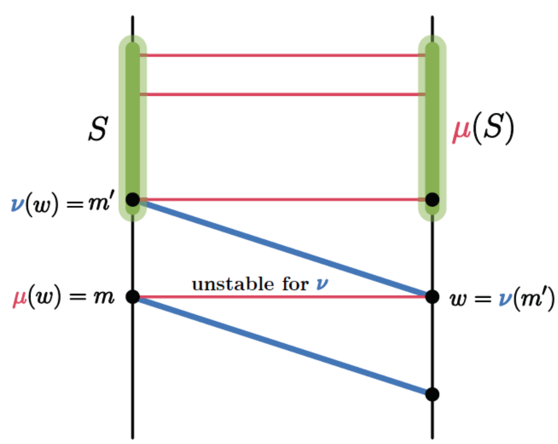


Figure 1: Illustration of scenario in Problem 1(b), where $\mu(S) \neq \nu(S)$. Points on the left and right vertical line represent proposers and proposees, respectively, and matches given by $\mu$ and $\nu$ are indicated by red and blue lines, respectively.