# Representing All Stable Matchings by Walking a Maximal Chain

Linda Cai
tcai@princeton.edu

Clayton Thomas
claytont@princeton.edu

October 14, 2019

**Abstract**

The seminal book of Gusfield and Irving [GI89] provides a compact and algorithmically useful way to represent the collection of stable matches corresponding to a given set of preferences. In this paper, we reinterpret the main results of [GI89], giving a new proof of the characterization which is able to bypass a lot of the "theory building" of the original works. We also provide a streamlined and efficient way to compute this representation. Our proofs and algorithms emphasize the connection to well-known properties of the deferred acceptance algorithm.

## 1 Introduction

Stable matching mechanisms are ubiquitous in theory and in practice, especially in the "bipartite case" where agents lie in two disjoint groups and one-to-one matches are made between members of different groups. The most commonly used stable matching mechanism is the Gale-Shapley algorithm, i.e. "one-side proposing deferred acceptance". This algorithm has the nice properties of being simple to implement, fast to execute, and strategyproof for the proposing side. However, deferred acceptance always returns the best stable matching for the proposing side and the worst stable match for the receiving side. This leads to a basic question: what lies in between?

This question can be rephrased as follows: how can one understand, represent, and traverse the set of all stable matchings for a given set of preference? An excellent answer to this question was given by [GI89], based on the works [IL86, Irv85, ILG87]. Despite the fact that there can be exponentially many stable matchings[1], the collection of *all* stable matching can be compactly represented in a form which is efficient to construct and algorithmically useful, and sheds light on the structure of the stable matching instance.

In this paper, we reinterpret and simplify the classification provided by [GI89]. We provide full proofs which characterize the "lattice structure" of the set of stable matchings and culminate in a theorem equivalent to the main characterization of [GI89]:

**Theorem 1.1** (Combination of theorems 4.9 and 5.5)**.** *For any stable matching instance, there is a directed acyclic graph $G$, computable in $O(n^2)$ time, such that there is a bijection between the set of all stable matchings and the collection of closed subsets of $G$ (i.e. the subsets $S$ of vertices of $G$ such that no directed edge $(u, v)$ of $G$ has $u \in S$ but $v \notin S$).*

There is a compelling interpretation of the vertices of $G$. They are called "rotations", and represent the fact that, starting from some stable matching, some set of men $(m_0, \ldots, m_{k-1})$ can "cyclically move partners" (i.e. each $m_i$ gets re-matched to the partner of $m_{i+1}$ (with indices mod

[1] For an easy example, consider $n/2$ "disjoint copies" of an instance with two men and two women which has two stable matchings. This has $2^{n/2}$ stable outcomes.

$k$)) to arrive at a new stable matching. A full description of the rotations, and the dependencies between them, is given in this paper. Indeed, the primary simplifying contribution of this paper is in focusing on rotations "from the start" instead of going through other notions.

We believe that our proof strategies and presentation is more intuitive and more "fundamentally algorithmic", as we focus on how simple properties of the ubiquitous deferred acceptance algorithm can lead us to understand the full set of stable matchings. Furthermore, we give a new perspective on the algorithm used in [GI89] to construct the compact representation. Along the way, we correct a minor error in the original algorithm from [GI89] (for details, see appendix B).

While the stable matching problem is a classic and well-studied problem, there are many exciting contemporary developments in the theory, from worst-case upper bounds on the number of stable matchings [KGW18] to the communication complexity of finding stable matchings [GNOR19] to detailed studies of different incentives properties [AG18, Gon14]. Our intent is for this paper to provide a starting point for researchers interested in studying stable matching markets from an algorithmic perspective.

### 1.1 Organization and relation to prior work

For completeness, we prove every result about stable matchings which we will need in this paper. In section 2, we make our formal definitions and review the basic properties of deferred acceptance and the set of stable matchings. Readers familiar with stable matchings can likely skip this section (possibly reviewing the lattice-theoretic vocabulary given in section 2.3). The core technical material is presented in sections 3 to 5.

- In section 3, we discuss how to traverse the stable matching lattice algorithmically. Intuitively, this involves women "rejecting" their current match, and continuing running deferred acceptance to get to a better stable matching (for the women). Our core technical tool, inspired by [IM05] is to use the concept of *deferred acceptance with truncated preferences*.
- In section 4, we define a compact representation of the stable matching lattice in terms of "minimal differences" called rotations, and prove that the representation is correct. Our definitions and theorems are as in [GI89], but we are able to significantly simplify our treatment by focusing on rotations "from the start" and avoiding intermediate representations. In particular, claim 4.7 and its proof using claim 4.6 are the key new ideas, which provide a way to show that a graph represents a lattice using a proof approach which (to the best of our knowledge) is brand new. Appendix A provides a detailed comparison of our methods and those of [GI89].
- In section 5, we show how to efficiently construct the compact representation defined in section 4. While our algorithm is essentially equivalent to that in [GI89] (figure 3.2 on page 110), we provide a more streamlined way to find the "predecessor relations" between rotations, which are the edges of the graph $G$, and thus avoid a minor error in the way that [GI89] finds these predecessor relations. In appendix B, we point out and correct this minor error. Our presentation is similar to that of the "MOSM to WOSM" algorithm in [AKL17], which relates more clearly to our conceptual use of deferred acceptance.

## 2 Stable Matchings and Deferred Acceptance

We start with the basic definitions. A matching market is a collection $\mathcal{M}$ of "men" and $\mathcal{W}$ of "women", where each man $m \in \mathcal{M}$ has a ranking over women in $\mathcal{W}$, represented as list ordered from most preferred to least preferred, and vice versa. Lists may be partial, and agents included on the list of some $a \in \mathcal{M} \cup \mathcal{W}$ are called the acceptable partners of $a$. We write $w_1 \succ_m w_2$ if $w_1$ is ranked higher than $w_2$ on $m$'s list (or if $w_1$ is acceptable but $w_2$ is not ranked at all). We also denote the fact that $w$ is not an acceptable partner of $m$ by $\emptyset \succ_m w$, and conversely if $w$ is an acceptable partner of $m$ we write $w \succ_m \emptyset$. A *matching* is a set of vertex disjoint edges in the bipartite graph $G(M, W)$, where $(m, w) \in E(G)$ if and only if $m$ is acceptable to $w$ and vice versa. We denote a matching by $\mu : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W} \cup \{\emptyset\}$, where $\mu(i)$ is the matched partner of agent $i$. We write $\mu(i) = \emptyset$ if agent $i$ is unmatched.

For a set of preferences $P = \{\succ_w\}_{w \in \mathcal{W}} \cup \{\succ_m\}_{m \in \mathcal{M}}$ and any matching $\mu$, a man/woman pair $(m, w)$ is called *blocking* if we simultaneously have $m \succ_w \mu(w)$ and $w \succ_m \mu(w)$. A matching $\mu$ is *stable* for a set of preferences $P$ if no unmatched man/woman pair is blocking for $P$. A *pair* $(m, w)$ is called stable for $P$ if $\mu(m) = w$ in *some* stable matching, and $m$ is called a stable partner of $w$ (and vice-versa).

## 2.1 $MPDA$ and the man-optimal stable matching

The most natural way to find stable matchings is with the celebrated *deferred acceptance algorithm.* In this paper, we consider man proposing deferred acceptance ($MPDA$) as given in Algorithm 1. For completeness, here we provide simple proofs of the basic but crucially important properties of this algorithm.

**Algorithm 1** $MPDA$: Men-proposing deferred acceptance

```
Let U = M be the set of unmatched men
Let μ be an all empty matching
while U ≠ ∅ and some m ∈ U has not proposed to every woman on his list do
    Pick such a m (in any order)
    m "proposes" to their highest-ranked woman w which they have not yet proposed to
    if m ≻_w μ(w) then
        If μ(w) ≠ ∅, add μ(w) to U
        Set μ(w) = m, remove m from U
```

Intuitively, this algorithm starts with the men doing whatever they prefer the most, then doing the minimal amount of work to make the matching stable. Indeed, men propose in their order of preference. If a woman $w$ ever rejected a man $m$ they prefer over their current match, then *remained* with their current match, then $(m, w)$ would clearly create an instability in the final matching.

**Claim 2.1.** *The output of $MPDA$ is a stable matching.*

*Proof.* First, observe that the $MPDA$ algorithm terminates because every man will propose to every woman at most once. The claim follows from two simple invariants of the algorithm:

- Men propose in their order of preference.
- Women can only increase the rank of their tentative match over time (and once they are matched, they stay matched).

Formally, consider a pair $m \in \mathcal{M}$, $w \in \mathcal{W}$ which is unmatched in the output matching $\mu$. Suppose for contradiction $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. In the $MPDA$ algorithm, $m$ would propose to $w$ before $\mu(m)$. This means that $w$ received a proposal from a man she preferred over her eventual match $\mu(w)$, a contradiction. □

Note that this algorithm gives us a very interesting existence result: it was not at all clear that stable matching existed before we had this algorithm.

We can now formalize our intuition that $MPDA$ does the least amount of work needed to result in a stable outcome (after men propose to their favorite women). We show that every rejection which happens in $MPDA$ *must* happen in order for a stable matching to result. The proof uses the following technique: although it's not immediately easy to show an event can't happen, you can show it *can't happen for the first time.*

**Claim 2.2.** *If a man $m \in \mathcal{M}$ is ever rejected by a woman $w \in \mathcal{W}$ during some run of $MPDA$ (that is, $m$ proposes to $w$ and $w$ does not accept) then no stable matching can pair $m$ to $w$.*

*Proof.* Let $\mu$ be any matching. Suppose that some pair, matched in $\mu$, is rejected during $MPDA$. Consider the first time during in the run of $MPDA$ where such a rejection occurs, i.e. a woman $w$ rejects $\mu(w)$ but no other woman $w'$ has rejected $\mu(w')$ so far. In particular, let $w$ reject $m = \mu(w)$ in favor of $m' \neq m$ (either because $m'$ proposed to $w$, or because $m'$ was already matched to $w$ and $m$ proposed). We have $m' \succ_w m$, so if $m'$ is unmatched in $\mu$, then $\mu$ is unstable. Thus we have $\mu(m') = w' \neq w$, and because this is the first time any man has been rejected by a match from $\mu$, $m'$ has not yet proposed to $w'$. Because men propose in their preference order, we have $w \succ_{m'} w'$. However, this means $\mu$ is not stable.

Thus, no woman can ever reject a stable partner in $MPDA$. $\square$

By the previous claim, $MPDA$ moves the men down their preference lists the minimal amount required to enforce stability. Interestingly, a completely dual phenomenon occurs for the women's preferences.

**Corollary 2.3.** *Let the set of men and women who receive a match at the end of $MPDA$ be denote $\mathcal{M}_{\text{matched}}$ and $\mathcal{W}_{\text{matched}}$, respectively. In this matching $\mu$:*

1. *every $m \in \mathcal{M}_{\text{matched}}$ is paired to his best stable match.*
2. *every $w \in \mathcal{W}_{\text{matched}}$ is paired to their worst stable match.*

*Proof.* Over the course of $MPDA$, each man $m \in \mathcal{M}_{\text{matched}}$ was rejected by every woman which he prefers to his partner in $MPDA$. By claim 2.2, this means his partner in $MPDA$ is his top stable match.

Let $m \in \mathcal{M}$ and $w \in \mathcal{W}$ be paired by $MPDA$. Let $\mu$ be any stable matching which does not pair $m$ and $w$. We must have $w \succ_m \mu(m)$, because $w$ is the best stable partner of $m$. If $m \succ_w \mu(w)$, then $\mu$ is not stable. Thus, $w$ cannot be stably matched to any man she prefers less than $m$. $\square$

The last claim also implies that the matching output by $MPDA$ is independent of the order in which men are selected to propose.

## 2.2 General stable matchings

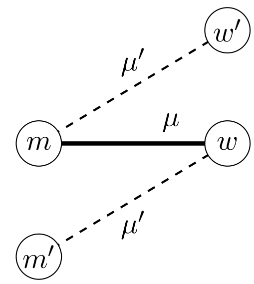
Interestingly, our claim 2.2, which related to $MPDA$, can be used to prove a fundamental property of the set of *all* stable matchings. Specifically, we can prove the following weaker version of the rural hospital theorem[2] which will be key for much of our discussion in section 3.

**Claim 2.4** (Rural Hospital Theorem)**.** *Then the set of unmatched agents is the same across every stable outcome.*

[2] The full rural hospital theorem [Rot86] applies to many-to-one matching markets (i.e. the residents and hospitals problem). The conclusion is that if a hospital does not fill *all* its openings in *some* stable outcome, then it will fail to fill all its openings (and indeed receive exactly the same doctors) in *every* stable outcome.

*Proof.* Let $\mathcal{M}_{\text{single}}$ be the set of men unmatched at the end of MPDA. Observe that each man in $\mathcal{M}_{\text{single}}$ has proposed to every acceptable partner he has over the run of MPDA. Thus, claim 2.2 implies that $\mathcal{M}_{\text{single}}$ is unmatched in every stable outcome. On the other hand, reversing the roll of men and women and considering women-proposing deferred acceptance, we can see that the set of (un)matched women is also identical across every stable outcome. □

Claim 2.3 seems to indicate that that the incentives of women and men are exactly opposite with regards to the results of man-proposing or women-proposing deferred acceptance. These next two claims prove that this is true for *all* stable matchings. In 2.3, we investigate these "order theoretic" properties further.




**Claim 2.5.** *Let $\mu, \mu'$ be stable matchings, and say $\mu(m) = w$, but $\mu'(m) \neq w$. Then $\mu'(m) \succ_m w$ if and only if $\mu'(w) \prec_w m$.*

*Proof.* ($\Leftarrow$) "If $w$ downgrades, then $m$ upgrades". Suppose $\mu'(w) \prec_w m$. Because $\mu'$ is stable, yet $m$ and $w$ are not matched in $\mu'$, we must have $\mu'(m) \succ_m w$, or else $(m, w)$ would form a blocking pair. (A rephrasing: this direction is easy because the definition of stability immediately makes it impossible for $m$ and $w$ to both downgrade).

($\Rightarrow$) "If $w$ upgrades, then $m$ downgrades". Let $m' = \mu'(w) \neq m$ and $w' = \mu'(m) \neq w$. Suppose that $m' \succ_w m$, and for contradiction suppose that $w' \succ_m w$. Because $\mu'$ is stable, $(m', w')$ is not a blocking pair, so either $w \succ_{m'} w'$ or $m \succ_{w'} m'$. In the first case, $(m', w)$ form a blocking pair in $\mu$, and in the second case, $(m, w')$ form a blocking pair in $\mu$. Thus, in either case $\mu$ is not stable. □

**Claim 2.6.** *Let $\mu$ and $\mu'$ be stable matchings. Every man (weakly) prefers their match in $\mu$ over $\mu'$ if and only if every woman (weakly) prefers their match in $\mu'$ over $\mu$.*

*Proof.* Suppose each $m \in \mathcal{M}$ has $\mu'(m) \succeq_m \mu(m)$. For each $w \in \mathcal{W}$ with $\mu(w) \neq \mu'(w)$, we must have $\mu'(w) \prec_w \mu(w)$ by claim 2.5. The proof for the other direction is identical. □

## 2.3 The Lattice of Stable Matchings

Given that men and women have strictly opposite incentives, it is natural to define a dominance relationship over all stable matchings according to the preferences of one side of the market.

**Definition 2.7.** *We say that a stable matching $\mu$* woman-dominates *$\mu'$, written $\mu \geq \mu'$, if for every $w \in \mathcal{W}$, we have $\mu(w) \succeq_w \mu'(w)$ (that is, every woman is at least as happy with her match in $\mu$ as in $\mu'$). For some fixed set of preferences, we let $\mathcal{L}$ denote the set of stable matchings of $P$, ordered by the relation $\geq$.*

We can define man-dominance analogously, and by claim 2.6, $\mu$ man-dominates $\mu'$ if and only if $\mu \leq \mu'$. Now, one can visualize the collection of all stable matchings as starting with the unique man-optimal outcome at the bottom, the unique woman-optimal outcome at the top, and all other stable matching in between.

In this section we show that the set of all stable matchings for a set of preferences $P$ forms what's called a *distributive lattice* under the women-dominance order. For the sake of completeness, we first discuss the relevant definitions. Informally, a lattice is a partial order in which, for any two elements $a, b$, there is a unique "lowest element above $a$ and $b$" (the join) and a "highest element below $a$ and $b$" (the meet)[3].

[3] Note that it follows from the definition that join and meet operations, if they exist, are unique.

**Definition 2.8.** *A* partial order $\leq$ *is a reflexive, transitive, antisymmetric relation. We write* $a < b$ *when* $a \leq b$ *and* $a \neq b$.

*For elements* $a, b$ *of a partial order, a* least upper bound $a \vee b$ *is an element such that* $a \leq a \vee b$ *and* $b \leq a \vee b$*, and for any* $c$ *such that* $a \leq c$ *and* $b \leq c$*, we have* $a \vee b \leq c$*. A* greatest lower bound $a \wedge b$ *is defined analogously, interchanging* $\leq$ *with* $\geq$*. We also call* $a \vee b$ *the* join *of* $a$ *and* $b$ *and* $a \wedge b$ *the* meet *of* $a$ *and* $b$.

*A* lattice $L$ *is a partial order in which there exist greatest lower bounds and least upper bounds for any* $a, b \in L$.

*A* chain *in a lattice is any "totally ordered" sequence* $a_1 \leq a_2 \leq \ldots \leq a_k$.

*A lattice* $L$ *is* distributive *if the join and meet operations satisfy the following equations:*

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$

The join and meet operations in $\mathcal{L}$ are very natural: the join of $\mu$ and $\mu'$ corresponds to the matching $\tilde{\mu}$ where each woman gets the better of her two partners from $\mu$ and $\mu'$. This is exactly the operation one would hope would work – clearly $\tilde{\mu}$ is the worst matching (for the women) in which women do at least as well as in $\mu$ and $\mu'$. We prove below that this operation always yields a stable matching.

**Definition 2.9.** *Given stable matchings* $\mu$ *and* $\mu'$*, define* $\mu \vee \mu'$ *such that, for each woman* $w$*,* $(\mu \vee \mu')(w)$ *is the most preferred partner of* $w$ *among* $\mu(w)$ *and* $\mu'(w)$*. Similarly, define* $\mu \wedge \mu'$ *such that each woman is matched to their least preferred partner from* $\mu$ *or* $\mu'$.

**Theorem 2.10.** *The collection* $\mathcal{L}$ *of all stable matchings of some instance form a distributive lattice under the dominance ordering* $\leq$*, with join and meet given by* $\vee$ *and* $\wedge$.

*Proof.* It's easy to see that $\leq$ forms a partial order on $\mathcal{L}$. We'll show that $\vee$ gives least upper bounds in $\mathcal{L}$. It's easy to see that, if $\tilde{\mu} = \mu \vee \mu'$ is a stable matching, then it must be the least upper bound of $\mu$ and $\mu'$.

First, we claim that $\tilde{\mu}$ is a matching. Suppose for contradiction that some man $m$ is the match of two women $w$ and $w'$ in $\tilde{\mu}$. Without loss of generality suppose $\mu(w) = m$, so $m = \mu(w) \succ_w \mu'(w)$, and $\mu'(w') = m$, so $m = \mu'(w') \succ_{w'} \mu(w')$. Applying claim 2.5 twice, we get that $w = \mu(m) \prec_m \mu'(m) = w'$ and also that $w' = \mu'(m) \prec_m \mu(m) = w$, a contradiction.

Second, we claim that $\tilde{\mu}$ is stable. Suppose that $(m, w)$ is a blocking pair for $\tilde{\mu}$, Certainly the partners of $m$ and $w$ must be from different matchings among $\mu$ or $\mu'$, say $\tilde{\mu}(m) = \mu'(m)$ and $\tilde{\mu}(w) = \mu(w) \neq \mu'(w)$. As $(m, w)$ is blocking, $w \succ_m \mu'(m)$ and $m \succ_w \mu(w)$. But by the definition of $\tilde{\mu}$, we have $\mu(w) \succ_w \mu'(w)$, so $m \succ_w \mu'(w)$ as well, and $\mu'$ is not stable.

Now we show that $\wedge$ gives the greatest lower bound in $\mathcal{L}$. By claim 2.5, this is equivalent to defining $\mu \wedge \mu'$ such that every man gets their best partner from $\mu$ or $\mu'$ (because $m = \mu(w) \prec_w \mu'(w)$ if and only if $w = \mu(m) \succ_m \mu'(m)$). Thus, the proof is identical to the proof given for $\vee$, interchanging men with women.

Finally, the join and meet operations are distributive for the same reason that the operations of min and max distribute over each other. In particular, we can fix a woman $w$ and see that (with max and min taken according to $\succ_w$)

$$\big(\mu_1 \wedge (\mu_2 \vee \mu_3)\big)(w) = \min\big\{\mu_1(w), \max\big\{\mu_2(w), \mu_3(w)\big\}\big\}$$

$$\max\big\{\min\big\{\mu_1(w), \mu_2(w)\big\}, \min\big\{\mu_1(w), \mu_3(w)\big\}\big\} = \big((\mu_1 \wedge \mu_2) \vee (\mu_1 \wedge \mu_3)\big)(w)$$

□

The most important lattice-theoretic concept we will need is the notion of *covering*. Informally, an element covers another in a lattice if there is no element between them in the ordering.

**Definition 2.11.** *For $a, b$ elements of a lattice, we say $a$* covers *$b$ when $a > b$ and no element $c$ exists with $a > c > b$.*

A useful equivalent definition of covering is the following: $a$ covers $b$ if and only if whenever $a \geq c > b$, we have $a = c$. Although the concept of covering relations is central to our paper, we need remarkably few formal properties of covering relations (or of lattices for that matter). Here is what we will need:

**Claim 2.12.** *In any finite lattice and for any $a \leq b$, there exists a sequence $a = a_0 < a_1 < \ldots < a_k = b$ (for some $k \geq 0$) such that $a_i$ covers $a_{i-1}$ for each $i$. Such a sequence is called a* maximal chain *between $a$ and $b$.*

*Proof.* If $a = b$ we are done. Otherwise, let $S$ be the set of all elements $c$ such that $a < c < b$, and induct on $|S|$. If $|S| = 0$, then $b$ covers $a$ and we are done. Otherwise, take any $c \in S$. Consider the set of all $d$ such that $a < d < c$. For such a $d$, we have $a < d < b$ and also $d \neq c$. Thus, there are strictly fewer than $|S|$ such $d$. Thus, by induction, there exists a maximal chain between $a$ and $c$. Similarly, there exists a maximal chain between $c$ and $d$, so the concatenation of these two chains gives us a maximal chain between $a$ and $b$.

□

We are interested in covering relations in $\mathcal{L}$ because they describe the "minimal differences" needed to go from one stable matching to another. The previous claim hints that one can describe any matching $\mu$ by giving the "covering relations leading up to $\mu$". Eventually, we will describe all covering relations (using "rotations") and show how you can represent all stable matchings as certain subsets of these "minimal differences".

# 3 Navigating the Lattice of Stable Matchings

In this section, we study how the lattice-theoretic properties of $\mathcal{L}$ start to manifest algorithmically in certain special cases of $MPDA$. We'll characterize the covering relations (and thus the entire structure of the lattice), essentially in terms of execution traces of $MPDA$. Intuitively, the main result is that, starting from any stable matching, if a woman has a better stable partner, then she can "divorce" her husband, and if we keep running $MPDA$, we will arrive at a stable matching preferred by that woman.

Consider a fixed set of input preferences $P$.

**Definition 3.1.** *For a set of preferences $P$ and matching $\mu$ stable under $P$, define $P(\mu)$ as follows: every woman $w$ matched in $\mu$ truncates the end of their preference list just after $\mu(w)$ (removing all men ranked worse than their current match), and every man $m$ matched in $\mu$ truncates the beginning their preference list just before $\mu(m)$ (removing all women ranked better than their current match). Women unmated in $\mu$ keep their full preference list, and men unmatched in $\mu$ are removed from $P(\mu)$.*

*For a woman $w$ matched in $\mu$, define $P_w(\mu)$ the same as $P(\mu)$ with one additional change: woman $w$ truncates her preference list one more place by removing her current match $\mu(m)$.*

Intuitively, $P(\mu)$ defines the state we are in after a deferred acceptance type algorithm reaching matching $\mu$: the women are still seeking to improve beyond their current match and the men are still proposing down their lists. On the other hand, $P_w(\mu)$ represents preferences corresponding to woman $w$ attempting to reach next to a better match than $\mu(w)$ (by rejecting $\mu(w)$ in $MPDA$).

In what follows, we call a match stable if it is stable for the original set of preferences $P$. If we need to refer to the fact that a match is stable for some truncated set of preferences $P_w(\mu)$, we will specify so. We denote $\mathcal{M}_{\text{matched}}$ and $\mathcal{M}_{\text{single}}$ as the set of men who are matched and unmatched respectively in the stable matchings with the original preferences $P$ (recall from 2.4 that these sets are uniquely determined). Define $\mathcal{W}_{\text{matched}}$ and $\mathcal{W}_{\text{single}}$ analogously.

Note that the execution of $MPDA(P_w(\mu_0))$ is quite a bit more simple than a general execution of $MPDA$. After the first proposal of each man in $\mathcal{M}_{\text{matched}}$ (i.e. each $m \in \mathcal{M}_{\text{matched}} \setminus \{\mu_0(w)\}$ proposes to and is accepted by $\mu_0(m)$), there is exactly one "free" man from $\mathcal{M}_{\text{matched}}$ at a time (i.e. one man who is not tentatively matched and still proposing down his list), until there is no longer a free man and the execution terminates. Specifically, the free man is initially $\mu_0(w)$, and if a proposal from the free man is accepted by a woman $w' \in \mathcal{W}_{\text{matched}} \setminus \{w\}$, the free man becomes $\mu_0(w')$. If a proposal is accepted by $w$ or a woman from $\mathcal{W}_{\text{single}}$, or if a man proposes to the last woman on his preference list, the algorithm terminates. In order to capture such an execution sequence we make the following definition:

**Definition 3.2.** *Given a stable matching $\mu_0$ and a woman $w \in \mathcal{W}_{\text{matched}}$, the* rejection chain of $w$ starting from $\mu_0$ *is the list $(w_1, m_1, w_2, m_2, \ldots, a_i)$ defined as follows:*

- *$w_1 = w$ and $m_1 = \mu_0(w)$*
- *The men $m_i$ are, in order, the men from $\mathcal{M}_{\text{matched}}$ which are free during the execution of $MPDA(P_w(\mu_0))$*
- *For each $i$, $w_{i+1}$ is the woman (if any) who accepts a proposal from $m_i$*
- *The list ends when the algorithm terminates*

*We also call such a list "the rejection chain of $MPDA(P_w(\mu_0))$" or just "the rejection chain" if $w$ and $\mu_0$ are understood.*

We start by establishing some basic properties relating rejection chains to the match returned by $MPDA(P_w(\mu_0))$. The proof is immediate.

**Claim 3.3.** *Let $\mu_0$ be a stable match and let $w \in \mathcal{W}_{\text{matched}}$. Let $\mu'$ be the result of $MPDA(P_w(\mu_0))$ and let $(w_1, m_1, w_2, m_2, \ldots, a_i)$ be the rejection chain of $w$ starting from $\mu_0$ (so $a_i$ denote the last agent in the rejection chain). Then exactly one of the following is true:*

- *$a_i \in \mathcal{W}_{\text{single}}$ is a woman who is now matched in $\mu'$*
- *$a_i$ is a man from $\mathcal{M}_{\text{matched}}$ who is now unmatched in $\mu'$*
- *$a_i = w$ (and $w$ receives a match in $\mu'$ if and only if $a_i = w$)*

*Moreover, the set of agents matched in $\mu'$ is the same as that in $\mu_0$ if and only if $a_i = w$.*

Our goal is to expore the stable matching lattice $\mathcal{L}$ using the operation $(\mu_0, w) \mapsto MPDA(P_w(\mu_0))$. Thus, the first thing we need to know is when this operation keeps us in the lattice $\mathcal{L}$ and when the result is an unstable matching.

**Claim 3.4.** *Let $\mu_0$ be stable and take $w \in \mathcal{W}_{\text{matched}}$. Let $MPDA(P_w(\mu_0))$ terminate in a matching $\mu'$. Then $\mu'$ is stable if and only if $w$ receives a match in $\mu'$.*

*Proof.* If $w$ is not matched in $\mu'$, then the set of matched agents differs between $\mu_0$ and $\mu'$. Thus $\mu'$ cannot possibly be stable by the rural hospital theorem 2.4.

On the other hand, suppose $w$ receives a match in $\mu'$. By the previous claim, this means that the set of agents matched in $\mu_0$ and $\mu'$ are identical. Consider how $MPDA(P_w(\mu_0))$ runs, converting from $\mu_0$ to $\mu'$. Observe that, because $w$ receives a match (which she prefers to $\mu_0(w)$), every woman can only improve their preference for their match.

For the sake of contradiction, suppose $(m', w')$ is a blocking pair in $\mu'$. Certainly, $\mu'$ is stable for the preferences $P_w(\mu_0)$. How can $(m', w')$ be a blocking pair in $P$ but not in $P_w(\mu_0)$? The only way is if one agent truncated the other off their preference list in $P_w(\mu_0)$. We have two cases:

1. Suppose $w'$ truncated $m'$. Then we have $m' \preceq_{w'} \mu_0(w') \preceq_{w'} \mu'(w')$. But $m' \succ_{w'} \mu'(w')$, a contradiction.
2. Now suppose $m'$ truncated $w'$. Then $w' \succ_{m'} \mu_0(m')$. But then $m' \succ_{w'} \mu'(w') \succeq_{w'} \mu_0(w')$, so $(m', w')$ are unstable in $\mu_0$, a contradiction.

□

Next, we need to know that, if we have not reached the woman-optimal stable match, then we can *always keep moving up in the lattice.* Intuitively, this is true because, whenever a stable matching exists, $MPDA$ will find it, so if a stable matching with women receiving good partners exists, then $MPDA$ will find it as well.

**Claim 3.5.** *If $\mu_0$ is a stable matching in which $w$ is not paired to her optimal stable partner, then $MPDA(P_w(\mu_0))$ will return a stable matching $\mu'$ which strictly woman-dominates $\mu_0$, i.e. $\mu' > \mu_0$.*

*Conversely, if $MPDA(P_w(\mu_0))$ fails to return a stable match, then $w$ is matched to her optimal stable partner in $\mu_0$.*

*Moreover, if $\mu'$ covers $\mu_0$ in $\mathcal{L}$, then $MPDA(P_w(\mu_0))$ returns $\mu'$ for* any *woman $w$ who receives a better partner in $\mu'$ than in $\mu_0$.*

*Proof.* Let $\mu^*$ be any stable matching in which $w$ has a better partner than in $\mu_0$, i.e. $\mu^*(w) \succ_w \mu_0(w)$. Without loss of generality, we can assume that $\mu^* \geq \mu_0$, because if this is not the case, we can replace $\mu^*$ with $\mu^* \vee \mu_0$. By the rural hospital theorem (claim 2.4), $\mu^*$ must have exactly the same set of matched agents as in $\mu_0$.

Let $\mu' = MPDA(P_w(\mu_0))$, and note that $\mu'$ is certainly stable for preferences $P_w(\mu_0)$. Because $\mu^* \geq \mu_0$ and $w$ gets matched strictly above $\mu_0(w)$, the matching $\mu^*$ is also stable for preferences $P_w(\mu_0)$. Thus, once again $\mu^*$ and $\mu_0$ have identical sets of matched agents. By claim 3.4, we conclude that $\mu'$ is stable for preferences $P$.

By the definition of $P_w(\mu_0)$, each woman will only accept a proposal in $MPDA(P_w(\mu_0))$ from a man she likes at least as much as in $\mu_0$. As $w$ receives a strictly better match in $\mu'$, we have $\mu' > \mu_0$.

For the converse, suppose $w$ is matched to her optimal stable partner in $\mu_0$. In $\mu = MPDA(P_w(\mu_0))$, $w$ will not accept a proposal except from a man ranked above her match in $\mu_0$. Thus, $\mu$ cannot possibly be stable, as then $w$ would be matched in $\mu$ to a stable partner better than $\mu_0(w)$.

Now, suppose $\mu'$ covers $\mu_0$, so that whenever $\mu' \geq \mu > \mu_0$, we have that $\mu = \mu'$, and let $w$ be any woman recieving a better match in $\mu'$ than in $\mu_0$. We claim that $\mu'$ is the man-optimal stable outcome in which each woman in $\mathcal{W}_{\text{matched}}$ receives a partner at least as good as in $\mu_0$, and where $w$ receives a strictly better partner. Indeed, if $\mu'$ were *not* this matching, then some $\mu$ would exist such that $\mu' > \mu > \mu_0$, and so $\mu'$ would not cover $\mu_0$. By the fact that MPDA returns the man-optimal stable outcome (claim 2.3) this exactly menas that $\mu'$ is the result of $MPDA(P_w(\mu_0))$.

□

The previous claims show that if a woman $w$ has a better stable match, she can reject her current match, and if we continue running deferred acceptance then $w$ will achieve a better outcome. Next we get a characterization of when these changes from matching to matching are as small as possible (i.e. when the new matching covers the old in the lattice $\mathcal{L}$).

**Claim 3.6.** *Suppose $\mu_0$ is stable and $MPDA(P_w(\mu_0))$ terminates in a stable matching $\mu' > \mu_0$. Then $\mu'$ covers $\mu_0$ if and only if during the run of $MPDA(P_w(\mu_0))$, no woman from $\mathcal{W}_{\text{matched}}$ receives more than one proposal from men who she strictly prefers to her match in $\mu_0$.*

*Proof.* For this proof, call a proposal *good* if it is made by some man $m'$ to some woman $w'$, where $w'$ prefers $m'$ to $\mu_0(w')$. Note that a woman does not necessarily accept a good proposal (if she has already seen a proposal from a man she likes even more).

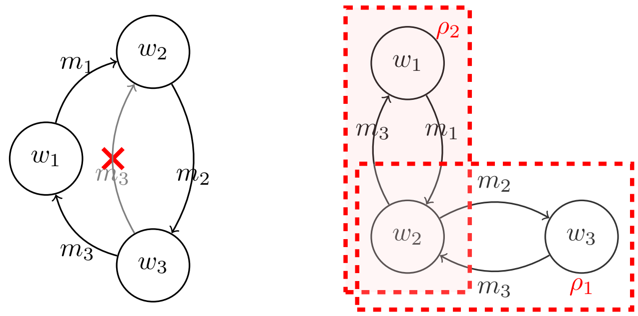
($\Rightarrow$) Suppose that, while running $MPDA(P_w(\mu_0))$, some woman sees more than one good proposal. Because $MPDA(P_w(\mu_0))$ terminates as soon as $w$ sees a good proposal, this woman cannot be $w$. Let $w^* \neq w$ be the *first* such woman, i.e. when $w^*$ receives her second good proposal, no other woman has yet received a second good proposal.

Consider running $MPDA(P_{w^*}(\mu_0))$, and call the result $\mu$. As $MPDA$ progresses, we know that each woman receives exactly one good proposal, because $w^*$ was the first instance where a woman received two good proposals. Thus, the rejection chain of $MPDA(P_{w^*}(\mu_0))$ is a sublist of the rejection chain of $MPDA(P_w(\mu_0))$, with one notable exception: $w^*$ might not accept her second good proposal in $MPDA(P_w(\mu_0))$, but she will definitely accept the corresponding proposal in $MPDA(P_{w^*}(\mu_0))$. Regardless of this event, every woman who changes partners in $MPDA(P_{w^*}(\mu_0))$ will also change partners in $MPDA(P_w(\mu_0))$, and indeed will do at least as well in the end, so $\mu \leq \mu'$. As $w$ could not have possibly changed partners in $MPDA(P_{w^*}(\mu_0))$, this means $\mu < \mu'$. We already knew that $\mu_0 < \mu$, so this completes the proof that $\mu'$ does not cover $\mu_0$.

($\Leftarrow$) For the other direction, suppose no woman sees multiple good proposals. Now suppose $\mu_0 < \mu \leq \mu'$ for some stable match $\mu$. Let $E = (w, m_1, w_2, m_2, \ldots, w_k, m_k, w)$ be the rejection chain of $MPDA(P_w(\mu_0))$. We'll show that, for any $i$, the outcome of $MPDA(P_{w_i}(\mu_0))$ is also $\mu'$. For each woman $w_i$ in $E$, consider the rejection chain $E_i$ of $w_i$ starting at $\mu_0$. For each man $m_j$, consider each woman $w$ on his preference list strictly between $w_j$ and $w_{j+1}$. Each such woman rejected him in $MPDA(P_w(\mu_0))$, and the only way for a woman to then accept $m_j$ in $MPDA(P_{w_i}(\mu_0))$ is if $m_j$ is a good proposal for $w$, but $w$ had already seen an (even better) good proposal in $MPDA(P_w(\mu_0))$. Because we assume no woman receives multiple good proposals, this is impossible, so each $w$ between $w_j$ and $w_{j+1}$ will still reject $m_j$. Furthermore, $w_{j+1}$ will still accept him, as she has not yet seen a good proposal when $m_j$ proposes to her. Thus, each link of $E_i$ will be the same as in $E$, that is, $E_i$ will simply be $(w_i, m_i, w_{i+1}, \ldots, m_{i-1}, w_i)$ (with indices taken mod $k$). Thus, the outcome of $MPDA(P_{w_i}(\mu_0))$ is $\mu'$.

Because $\mu_0 \neq \mu$, some woman must receive a strictly better match in $\mu$ than in $\mu_0$. As $\mu \leq \mu'$, that woman must be $w_i$ for some $i$. Because $MPDA$ returns man-optimal stable outcomes (claim 2.3), $\mu' = MPDA(P_{w_i}(\mu_0))$ is the man-optimal stable outcome in which every woman receives a match at least as good as in $\mu_0$, and in which $w_i$ receives a strictly better match. As $\mu$ is such a matching, we have $\mu' \leq \mu$ and thus $\mu' = \mu$. Because $\mu'$ was an arbitrary element of $\mathcal{L}$ with $\mu_0 < \mu \leq \mu'$, we've shown that $\mu'$ covers $\mu_0$. □

**Remark.** With the results of this section, we could already build the entire stable matching lattice $\mathcal{L}$, represented by its covering relations. Namely, we could essentially breadth-first search the lattice $\mathcal{L}$, finding those matching which cover a given $\mu_0$ by calculating $MPDA(P_w(\mu_0))$ for each woman $w$ (and keeping track of whether any woman receives multiple proposals from a man she prefers to her old match from $\mu_0$).




| | | | |
|---|---|---|---|
| $m_1$ | $w_1$ | $w_2$ | |
| $m_2$ | $w_2$ | $w_3$ | |
| $m_3$ | $w_3$ | $w_2$ | $w_1$ |

| | | | |
|---|---|---|---|
| $w_1$ | $m_3$ | $m_1$ | |
| $w_2$ | $m_1$ | $m_3$ | $m_2$ |
| $w_3$ | $m_2$ | $m_3$ | |

Figure 1: A rejection chain with no repeated agents does not imply a covering relationship

**Example.** The condition in claim 3.6 is subtly different from an agent appearing multiple times in the rejection chain. For instance, consider the example illustrated in Figure 1, and let $\mu_0 = \{(m_1, w_1), (m_2, w_2), (m_3, w_3)\}$ be the man-optimal stable outcome. Agents only appear one time in $MPDA(P_{w_1}(\mu_0))$'s rejection chain $(w_1, m_1, w_2, m_2, w_3, m_3, w_1)$. However, $\mu^* := \{(m_1, w_2), (m_2, w_3), (m_3, w_1)\}$, the resulting stable matching from $MPDA(P_{w_1}(\mu_0))$, does not cover $\mu_0$. In fact, $MPDA(P_{w_2}(\mu_0))$ results in a stable matching $\mu_1 = \{(m_1, w_1), (m_2, w_3), (m_3, w_1)\}$ such that $\mu_0 < \mu_1 < \mu^*$. Intuitively, what happened is that, because $w_2$ received a proposal from both $m_1$ and $m_3$ in $MPDA(P_{w_1}(\mu_0))$, she actually had *two* opportunities to upgrade and reach a better stable matching. Thus, the change from $\mu_0$ to $\mu^*$ can be broken down into two steps, where one step must come before the other (namely, $\mu_1$ must be reached before $\mu^*$). In the next section, we'll see how to formalize these concepts using *rotations* and *predecessor relations*.

# 4 Rotations

We now define a concise way to describe the difference between "consecutive" stable matchings, i.e. pairs of matchings where one covers the other. The collection of these "minimal differences" will allow us to represent all stable matchings in a principled and compact way.

**Definition 4.1.** *Let $\mu \in \mathcal{L}$ be a stable matching and $\rho = [(w_0, m_0), (w_1, m_1), \ldots, (w_{k-1}, m_{k-1})]$ a list of agents with each $w_i \in \mathcal{W}$ and $m_i \in \mathcal{M}$, and $\mu(w_i) = m_i$ for each $i$. The* elimination of $\rho$ from $\mu$ *is the matching $\mu'$ such that $\mu'(m_i) = w_{i+1}$ for $i = 0, \ldots, k-1$ (with indices taken mod $k$) and $\mu'(m) = \mu(m)$ for each $m$ which doesn't appear in $\rho$.*

*We say $\rho$ is a* rotation exposed in $\mu$ *when $\mu'$ is a stable matching, and $\mu'$* ***covers*** *$\mu$.*

*The collection of all rotations which are exposed in some $\mu \in \mathcal{L}$ is called the* set of rotations*, and is denoted by $\Pi$.*

We can "visualize" rotations as follows: if the men in the rotation all "get up" from their match in $\mu$ and move one place to the right (cyclically) in the rotation, then we arrive at a new stable match (which covers the old one). Note that we only call a list of agents $\rho$ a rotation when there exists a $\mu_0$ such that the elimination of $\rho$ from $\mu_0$ covers $\mu_0$. We view two rotations as equivalent if they differ by a cyclic shift, i.e. $\rho$ as above is identified with $[(w_i, m_i), (w_{i+1}, m_{i+1}), \ldots, (w_{i-1}, m_{i-1})]$ (indices taken mod $k$) for any $i$. By the definition, it is clear that such a shift changes nothing. For the rest of this section, all indices in rotations are considered mod $k$ where $k$ is the length of the rotation. We say each pair $(w_i, m_i)$ *appears* in rotation $\rho$, and that $\rho$ *moves* $m_i$ *from* $w_i$ *to* $w_{i+1}$ and *moves* $w_i$ *from* $m_i$ *to* $m_{i-1}$. If $\rho$ moves $m$ from $w_i$ to $w_{i+1}$, and $m$ ranks $w$ between $w_i$ and $w_{i+1}$ (that is, $w_i \succ_m w \succ_m w_{i+1}$), we say that $\rho$ moves $m$ from above $w$ to below $w$. Define the meaning of the phrase "$\rho$ moves woman $w$ from below $m$ to above $m$" and related phrases analogously.

Given the discussion in the previous section, we arrive easily at a rich set of claims characterizing rotations and their relationship to each other. Claim 4.2 translates the language of rotations to the concept of $MPDA$ with truncated lists, as discussed in section 3, then claim 4.3 lists the basic properties of rotations.

**Claim 4.2.** *For any stable matching $\mu_0$, the following are equivalent:*

- $\rho = [(w_0, m_0), \ldots, (w_{k-1}, m_{k-1})]$ *is a rotation exposed in $\mu_0$, and $\mu'$ is the elimination of $\rho$ from $\mu_0$*
- $MPDA(P_{w_0}(\mu_0))$ *produces the stable matching $\mu'$, and during its execution no woman receives multiple proposals from a man she prefers to her match in $\mu_0$, and the rejection chain of $w_0$ starting from $\mu_0$ is exactly $(w_0, m_0, w_1, m_1, \ldots, w_{k-1}, m_{k-1}, w_0)$.*

*Moreover, $\mu'$ covers $\mu_0$ if and only if there exists a rotation $\rho$ exposed in $\mu_0$ such that $\mu'$ is the elimination of $\rho$ from $\mu_0$.*

*Proof.* By claim 3.6, $\mu' = MPDA(P_{w_0}(\mu_0))$ is a stable matching which covers $\mu_0$ if and only if during its execution no woman receives multiple proposals from a man she prefers to her match in $\mu_0$, and $w_0$ receives a match in $\mu'$. In this case, the rejection chain is of the form $(w_0, m_0, w_1, m_1, \ldots, w_{k-1}, m_{k-1}, w_0)$, and the stable matching $\mu'$ is exactly is exactly the elimination of $\rho = [(w_0, m_0), \ldots, (w_{k-1}, m_{k-1})]$ from $\mu_0$.

By definition, the elimination of a rotation from $\mu_0$ always covers $\mu_0$. Furthermore, claim 3.5 tells us that whenever $\mu'$ covers $\mu_0$, running $MPDA(P_w(\mu_0))$ will produce $\mu'$ (and by claim 3.6 this will differ from $\mu_0$ by the elimination of a rotation). □

**Claim 4.3.** *We have the following:*

1. *$(w, m)$ are stable partners (i.e. matched in some stable matching) if and only if $(w, m)$ appears in some rotation in $\Pi$ or $(w, m)$ are paired in the woman-optimal stable outcome.*
2. *Let $(w_i, m_i)$ appear in some rotation (indexed as above). Then $m_{i-1}$ is the worst-ranked stable partner of $w_i$ who $w_i$ ranks above $m_i$ (and similarly $w_{i+1}$ is the best-ranked stable partner of $m_i$ who $m_i$ ranks below $w_i$). In other words, rotations move agents to their "next" stable partners (for women, the next best stable partner, and for men, the next-worst).*
3. *A pair $(w, m)$ of men and women appear in at most one rotation together.*
4. *There are at most $\binom{n}{2}$ rotations in $\Pi$.*

*Proof.* (1) The "if" direction is true by definition. For the "only if" part, let $\mu_0$ be a matching other than the woman-optimal outcome, and let $\mu_0(m) = w$ for $(m, w)$ not paired in the woman-optimal outcome. Let $\mu'$ be the woman-optimal stable outcome. Consider any maximal chain $\mu_0 < \mu_1 < \ldots < \mu_k = \mu'$ between $\mu_0$ and $\mu_k$ (i.e. $\mu_i$ covers $\mu_{i-1}$ for each $i$). Because $w$ is not matched to $m$ in $\mu'$, there must be some covering relation $\mu_{i-1} < \mu_i$ where $w$ is at $m$ in $\mu_{i-1}$ but not in $\mu_i$. By claim 4.2, this corresponds to a rotation in which $(m, w)$ appears.

(2) Let the rotation $\rho$ be exposed in $\mu_0$ and let the elimination of $\rho$ from $\mu_0$ be $\mu'$. Suppose for the sake of contradiction that $w_i$ has a stable partner $m^*$ who she ranks between $m_{i-1}$ and $m_i$, i.e. $m_i \prec_w m^* \prec_w m_{i-1}$. Let $\mu^*$ pair $w_i$ and $m^*$. Consider the matching $\mu = (\mu_0 \vee \mu^*) \wedge \mu'$. We have $\mu \le \mu'$, and because $\mu_0 \le \mu'$ and $\mu_0 \le \mu_0 \vee \mu^*$, we also get $\mu_0 \le \mu$. Because $\mu(w_i) = m^*$, that means $\mu_0 < \mu < \mu'$, which contradicts the fact that $\mu'$ covers $\mu_0$. This proves that $m_{i-1}$ is the worse stable partner of $w_i$ after $m_i$. The proof that, for each man $m_i$, $w_{i+1}$ is the next best stable partner $m_i$ has below $w_i$ is analagous.

(3) By part 2, given one pair $(m_i, w_i)$, the value of $w_{i+1}$ is uniquely determined as the next stable partner of $m_i$ below $w_i$. But then, by considering $m_i$ and $w_{i+1}$, the value of $m_{i+1}$ is uniquely

determined as the worst stable partner of $w_{i+1}$ before $m_i$. Continuing this process, we see that specifying one pair in $\rho$ uniquely determines all of $\rho$.

(4). By part 1, each man can appear with at most $n-1$ agents in some rotation (each of his stable partners except his partner in the woman-optimal outcome), and by part 3 that pair can appear at most once. At least two pairs of agents appear in each rotation, so the total number of rotations is at most $n(n-1)/2$.

□

We already know a decent amount about the structure of individual rotations. However, the rotations interact in a specified way. In particular, there is a natural *ordering* among them – some rotations must be eliminated before others. To make this precise, we need some definitions.

**Definition 4.4.** *Let $\rho_1$ and $\rho_2$ be rotations in $\Pi$.*

1. *If there exists a man/woman pair $(w, m)$ such that $\rho_1$ moves $m$ to $w$, and $(w, m)$ appear in $\rho_2$ (i.e. $\rho_2$ moves $m$ away from $w$), then $\rho_1$ is called a* type 1 predecessor *of $\rho_2$*
2. *If there exist a man/woman pair $(w, m)$ such that:*
   - *$\rho_1$ moves $w$ from $m_j$ to $m_{j-1}$, and $m_j \prec_w m \prec_w m_{j-1}$ ($\rho_1$ moves $w$ from below to above $m$)*
   - *$\rho_2$ moves $m$ from $w_i$ to $w_{i+1}$, and $w_i \succ_m w \succ_m w_{i+1}$ ($\rho_2$ moves $m$ from above to below $w$)*

   *Then $\rho_1$ is called a* type 2 predecessor *of $\rho_2$.*

*If $\rho_1$ is either a type 1 or type 2 predecessor of $\rho_2$, we say $\rho_1$ is a predecessor of $\rho_2$.*

*Define the predecessor graph $G(\Pi)$ on rotations as follows: the vertices are every rotation $\rho \in \Pi$, for every $\rho_1, \rho_2$ such that $\rho_1$ is a predecessor of $\rho_2$, there is a directed edge from $\rho_1$ to $\rho_2$ (labeled according to whether they are predecessors of type 1 or type 2 (or both)).*

In short, $\rho_1$ is a type 1 predecessor of $\rho_2$ if $\rho_1$ move a couple $(m, w)$ together who $\rho_2$ moves apart. In this case, $\rho_1$ must be eliminated first by definition. Intuitively, $\rho_1$ is a type 2 predecessor of $\rho_2$ if running $MPDA$ to eliminate $\rho_2$ would trigger the elimination of $\rho_1$. Otherwise, for the pair $(m, w)$ in the definition, $m$ would propose to $w$ as he moves from $w_i$ to $w_{i+1}$, and $w$ would accept that proposal and trigger the elimination of $\rho_1$ (finally giving $w$ an even better match than $m$).

It turns out that the above two types of predecessor relations are necessary and sufficient to characterize which rotations must be eliminated before each other. More precisely, a permutation of the set of all rotations can be eliminated, one after the other, if and only if they are topologically sorted in the graph $G(\Pi)$[4].

We first prove that every possible sequence of eliminations forms a topological sort of $G(\Pi)$. As we will formally spell out in 4.9, this means that topological sorts of $G(\Pi)$ suffice to represent all stable matchings.

**Claim 4.5.** *Consider any chain $\mu_0 < \mu_1 < \ldots < \mu_k$ in $\mathcal{L}$ where $\mu_0$ is man-optimal, $\mu_k$ is woman-optimal, and $\mu_{i+1}$ covers $\mu_i$ for each $i$ (i.e. consider a maximal chain in $\mathcal{L}$). Then*

1. *Then there exists a unique sequence of rotations $\rho_0, \rho_1, \ldots, \rho_{k-1} \in \Pi$ such that $\mu_{i+1}$ is the elimination of $\rho_i$ from $\mu_i$ for each $i$.*
2. *Every rotation in $\Pi$ appears exactly once in this sequence.*

[4] A topological sort of a directed acyclic graph is a permutation of the vertices of the graph such that, for each directed edge $(u, v)$ in the graph, $u$ comes before $v$.

*3. If $\rho$ is a predecessor of $\rho^*$ (of type 1 or type 2), then $\rho$ appears before $\rho^*$ in this sequence.*

*Proof.* (1) By claim 4.2, such a $\rho_i$ exists for each $i$. Furthermore, given $\mu_i$ and $\mu_{i+1}$, it's clear that $\rho_i$ is uniquely determined.

(2) We showed in claim 4.3, part 2 that rotations (and hence covering relations) move agents up or down one place on their list of stable partners. Thus, over the course of the maximal chain, every stable pair must be matched in some $\mu_i$ (or else those agents could not reach their match in the woman-optimal outcome). Moreover, each stable pair which is not matched in $\mu_k$ must appear in some rotation $\rho_i$. Because a stable pair appears in at most one rotation (claim 4.3, part 3), this means every rotation in $\Pi$ is in this sequence.

(3) Assume for contradiction that $\rho_j$ is a predecessor of $\rho_i$ for $i < j$. We have two cases.

Suppose $\rho_j$ is a type 1 predecessor of $\rho_i$. By definition, there exists a pair $(m, w)$ such that $\rho_j$ moves $m$ to $w$, and $(m, w)$ appears in $\rho_i$. By 4.3, part 2, rotations always move women to men which they rank higher than their current match, this means that in $\mu_j$, $w$ was matched below $m$ (i.e. $m \succ_w \mu_j(w)$). But $(m, w)$ are matched in $\mu_i$. Thus, we cannot have $\mu_i \le \mu_j$, a contradiction.

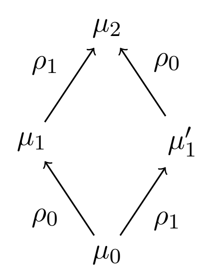
Now suppose $\rho_j$ is a type 2 predecessor of $\rho_i$. By definition, there exist $(m, w)$ such that $\rho_j$ moves $w$ from below $m$ to above $m$ and $\rho_i$ moves $m$ from above $w$ to below $w$. Thus, in $\mu_j$, $w$ is matched below $m$, and in $\mu_{i+1}$, $m$ is matched below $w$. Now, $\mu_{i+1} \le \mu_j$, so in $\mu_j$, $w$ is also matched below $m$. But this means that $\mu_{i+1}$ is not stable, a contradiction. □

Note that part 3 above implies that $G(\Pi)$ is an acyclic graph.

We'll see next that *only* stable matchings arise in the way described by the previous claim. In other words, the type 1 and type 2 predecessor relations are the *only* issues to applying any sequence of rotations you would like.

Our strategy will be to show that every topological sort of $G(\Pi)$ corresponds to a maximal chain in $\mathcal{L}$. We start with an arbitrary chain which corresponds to some topological sort, and the apply a special type of "commutativity operation" in order to transform that initial topological sort into the one we want (while preserving the property of corresponding to some maximal chain along the way). The next claim is a technical lemma which provides the type of commutativity operation needed. Another way to summarize this claim is that two adjacent rotations which are not predecessors do not interfere with each other.

**Claim 4.6.** *Suppose $\mu_0 < \mu_1 < \mu_2$, where each matching covers the previous one. Let $\rho_0$ and $\rho_1$ be the corresponding rotations, i.e. $\mu_{i+1}$ is the elimination of $\rho_i$ from $\mu_i$ for $i = 0, 1$. Assume that $\rho_0$ is not a predecessor of $\rho_1$ in $G(\Pi)$. Then $\rho_1$ is also exposed in $\mu_0$. Moreover, if $\mu_1'$ is the elimination of $\rho_1$ from $\mu_0$, then $\rho_0$ is exposed in $\mu_1'$, and $\mu_2$ is the elimination of $\rho_0$ from $\mu_1'$.*




*Proof.* Recall that every agent who does not appear in $\rho_0$ receives the same match in $\mu_0$ and $\mu_1$.

First, we claim that no agent can appear in both rotations. Proof: If some man $m$ appeared in both rotation, then $\rho_0$ moved $m$ to some woman $\mu_1(m)$, and then $(\mu_1(m), m)$ must appear in $\rho_1$. If a woman $w$ appears in both rotations, then the pair $(w, \mu_1(w))$ appears in $\rho_1$, so $\rho_0$ must move $\mu_1(w)$ to $w$. Because we've assumed that $\rho_0$ is not a type 1 predecessor of $\rho_1$, neither of the above cases can occur.

Now, let $w_0$ be a woman appearing in $\rho_1$ and consider $MPDA(P_{w_0}(\mu_0))$. We claim that this produces a stable outcome and no woman receives more than one proposal from a man she prefers to her match in $\mu_0$. Proof: Let $\rho_1 = [(w_0, m_0), (w_1, m_1), \ldots, (w_{k-1}, m_{k-1})]$. In $MPDA(P_{w_0}(\mu_0))$, the free man is initially $m_0$. Consider the proposals that $m_i$ makes after he is rejected by $w_i$. Some of the women he proposes to may be at different matches in $\mu_0$ than in $\mu_1$ (specifically, those women who were moved by $\rho_1$). However, for all such women $w$ who $m_i$ ranks above $w_{i+1}$, $w$ cannot be matched below $m_i$, or else $\rho_0$ would be a type 2 predecessor of $\rho_1$ by definition. Because none of the agents in $\rho_2$ appear in $\rho_1$, $w_{i+1}$ has the same match in $\mu_0$ as in $\mu_1$. Thus, all the women $m_i$ proposes to before $w_{i+1}$ will reject him, but $w_{i+1}$ will accept him. Thus, by induction, the rejection chain of $MPDA(P_{w_0}(\mu_0))$ will be exactly the same as in $MPDA(P_{w_0}(\mu_1))$. By claim 4.2, this means $\rho_1$ is exposed in $\mu_0$ and that and $\mu'_1 = MPDA(P_{w_0}(\mu_0))$ is the elimination of $\rho_1$ in $\mu_0$.

Finally, consider running $MPDA(P_{w'_0}(\mu'_1))$ for some $w'_0$ appearing in $\rho_0$. Let $\rho_1 = [(w'_0, m'_0), \ldots, (w'_{k'-1}, m'_{k'-1})]$, and again consider a free man $m'_i$ during this rejection chain. The only difference between $\mu'_1$ and $\mu_0$ is that the women who appear in $\rho_1$ have received better partners. However, $w_{i+1}$ is still matched to $\mu_0(w_{i+1})$, again because the agents in $\rho_0$ and $\rho_1$ are disjoint. The women $m_i$ proposes to before $w_{i+1}$ can only have higher matches than in $\mu_0$, so they will still reject his proposals. But $w_{i+1}$ will still accept. Thus, $MPDA(P_{w^*}(\mu'_1))$ will terminate with exactly $\rho_1$ eliminated from $\mu'_1$, and no woman will recieve multiple proposals from a man she prefers to her match in $\mu'_1$. So $\rho_1$ was exposed in $\mu'_1$.

Finally, it's clear from the definitions that the elimination of $\rho_0$ from $\mu'_1$ is $\mu_2$ (the match of every agent is uniquely determined as either the match from $\mu_0$ or the match which is uniquely specified in $\rho_0$ or $\rho_1$). □

Now we can prove that only maximal chains arise from topological sorts of $G(\Pi)$. One short way to summarize this proof is the following: we can transform any two topological sorts of $G(\Pi)$ between each other using only the "adjacent order swapping" operation given by the previous lemma[5]. Thus, starting from a fixed topological sort of $G(\Pi)$ (which corresponds to a maximal chain by claim 4.5) we see than any other topological sort will also correspond to a maximal chain.

**Claim 4.7.** *Consider any topological sort $\rho_0, \rho_1, \ldots, \rho_{k-1}$ of $G(\Pi)$, i.e. an ordering of each element of $\Pi$ such that whenever $\rho_i$ is a predecessor of $\rho_j$, we have $i < j$. Then this sequence corresponds to a maximal chain $\mu_0, \mu_1, \ldots, \mu_k$ in the stable matching lattice $\mathcal{L}$ such that $\mu_0$ is the man-optimal stable outcome, $\mu_{i+1}$ is the elimination of $\rho_i$ from $\mu_i$, and $\mu_k$ is the woman-optimal outcome.*

*Proof.* For this proof, say that a permutation $\rho'_0, \ldots, \rho'_{k-1}$ of the set $\Pi$ is *valid* if there exists $\mu'_0, \mu'_1, \ldots, \mu'_k$ a maximal chain in $\mathcal{L}$ such that $\mu_{i+1}$ is the elimination of $\rho_i$ from $\mu_i$ for each $i$.

Fix any arbitrary maximal chain $\mu_0 < \mu_1 < \ldots < \mu_k$ in $\mathcal{L}$. By claim 4.5, there exists a corresponding valid sequence $\rho_0, \rho_1, \ldots, \rho_{k-1}$ of rotations which is a topological sort of $G(\Pi)$. Now, given any permutation of $0, 1, \ldots, k-1$, say say $i_0, i_1, \ldots, i_{k-1}$, we'll prove that $\rho_{i_0}, \rho_{i_1}, \ldots, \rho_{i_{k-1}}$ is valid using induction on the *number of inversions* of the permutation (i.e. the number of pairs $j < k$ such that $i_j > i_k$). If there are no inversions, then $i_j = j$ for each $j$ and we are done.

Now, suppose $I = i_0, i_1, \ldots, i_{k-1}$ has at least one inversion and let $i_j, i_{j+1}$ be any *adjacent* inverted pair (if no adjacent inverted pairs exist, then no inverted pairs can exists). Consider the ordering $I' = i_0, \ldots, i_{j-1}, i_{j+1}, i_j, i_{j+2}, \ldots, i_{k-1}$. There is exactly one fewer inverted pair in this new ordering than in the original one, so by induction the ordering on rotations corresponding to $I'$ is valid. Say this ordering on rotations corresponds to a maximal chain containing the matchings $\mu_a < \mu_b < \mu_c$, where $\mu_b$ is the elimination of $\rho_{i_{j+1}}$ from $\mu_a$ and $\mu_c$ is the elimination of $\rho_{i_j}$ from $\mu_b$. Because both $\rho_0, \ldots, \rho_{k-1}$ and $\rho_{i_0}, \ldots, \rho_{i_{k-1}}$ are topological sorts of $G(\Pi)$, $\rho_{i_j}$ and $\rho_{i_{j+1}}$ cannot be in a predecessor relation. Thus, applying claim 4.6 to the covering relations $\mu_a < \mu_b < \mu_c$, we see that the original ordering $\rho_{i_0}, \rho_{i_1}, \ldots, \rho_{i_{k-1}}$ is also valid (and only the matching $\mu_b$ is different along the corresponding maximal chain).

[5] One way to do this (different then outlined in our formal proof) is to label the first list of rotations with $0, 1, \ldots, k-1$, then simply *bubble sort* the second list of rotations.

Thus, every topological sort of $G(\Pi)$ is valid by induction.

□

Finally, after one more simple definition, we arrive at our long sought after bijection.

**Definition 4.8.** *A* closed *subset $S$ of $G(\Pi)$ is a collection of rotations such that, whenever $\rho_2$ is in $S$ and $\rho_1$ is a predecessor of $\rho_2$, then $\rho_1$ is in $S$.*

**Theorem 4.9.** *There is a bijection between the collection of closed subsets of $G(\Pi)$ and the stable matching lattice $\mathcal{L}$.*

*This bijection is given as follows: for a closed subset $S$ of $\Pi$, let $\rho_0, \ldots, \rho_i$ be a topological sort of $S$ in $G(\Pi)$. Let $\mu_0$ be the man-optimal stable outcome, and let $\mu_{j+1}$ be the elimination of $\rho_j$ from $\mu_j$ for each $j = 0, \ldots, i$. Then the matching corresponding to $S$ is given by $\mu_{i+1}$.*

*Furthermore, let $\mu_1$ and $\mu_2$ be stable matchings corresponding to $S_1$ and $S_2$ respectively. Then $\mu_2$ woman-dominates $\mu_1$ (i.e. $\mu_2 \geq \mu_1$) if and only if $S_2 \supseteq S_1$.*

*Proof.* First, we show that this correspondence is surjective. Given a matching $\mu$, consider a maximal chain $\mu_0 < \mu_1 < \ldots < \mu_k$ containing it, say $\mu = \mu_i$. By claim 4.5, there exists a corresponding sequence of rotations $\rho_0, \rho_1, \ldots, \rho_{k-1}$. Consider the set $S = \{\rho_j\}_{j<i}$. For each $\rho_j \in S$ and $\rho'$ a predecessor of $\rho$, $\rho'$ must also be in $S$ because $\rho_0, \rho_1, \ldots, \rho_{k-1}$ is a topological sort. So $S$ is closed in $G(\Pi)$. Furthermore, $\mu = \mu_i$ is exactly given by the successive elimination of the rotations $\rho_0, \ldots, \rho_{i-1}$, starting from $\mu_0$. So $S$ corresponds to $\mu$.

Next, we show the correspondence is injective. Let $S_1, S_2$ be distinct closed subsets of $G(\Pi)$, let matching $\mu_1, \mu_2$ correspond to $S_1, S_2$, and without loss of generality take $\rho \in S_1 \setminus S_2$. Because $\rho$ has been eliminated in $\mu_1$ but not in $\mu_2$, any woman $w$ appearing in $\rho$ must prefer their match in $\mu_1$ to their match in $\mu_2$. Thus, $\mu_1$ and $\mu_2$ cannot be the same matching.

Finally, let $\mu_1, \mu_2 \in \mathcal{L}$ correspond to $S_1, S_2 \subseteq G(\Pi)$. We have $\mu_1 \leq \mu_2$ if and only if each woman $w$ does at least as well in $\mu_2$ as in $\mu_1$. For a fixed woman $w$, this occurs if and only if every rotation involving $w$ which appears in $S_1$ also appears in $S_2$. This is equivalent to the condition that every rotation which appears in $S_1$ also appearing in $S_2$, i.e. $S_1 \subseteq S_2$.

□

**Remark:** Because the bijection above respects ordering (i.e. $\mu_2 \geq \mu_1$ if and only if $S_2 \supseteq S_1$), the bijection above is actually a *lattice isomorphism.* So joins and meets in $\mathcal{L}$ correspond to joins and meets in the lattice of close subsets of the graph $G(\Pi)$, which is given by set union and set intersection, respectively.

# 5 Efficiently Finding Rotations

## 5.1 A simple example

We start with an example of how to use the facts proven above. Let men and women's preference list be as illustrated in the right column of Figure 2. The borders in the table highlight the preferences which cause the different rotations to form, and the boldfaced entries correspond to a type 2 predecessor relationship.

For the sake of illustration, let us denote a stable matching $\mu$ as $(\mu(m_1)\mu(m_2)...\mu(m_n))$ (e.g. (123456) means every $m_i$ is matched to $w_i$). It is easy to see that $\mu_0 = (123456)$ is the man-optimal stable matching in this example.

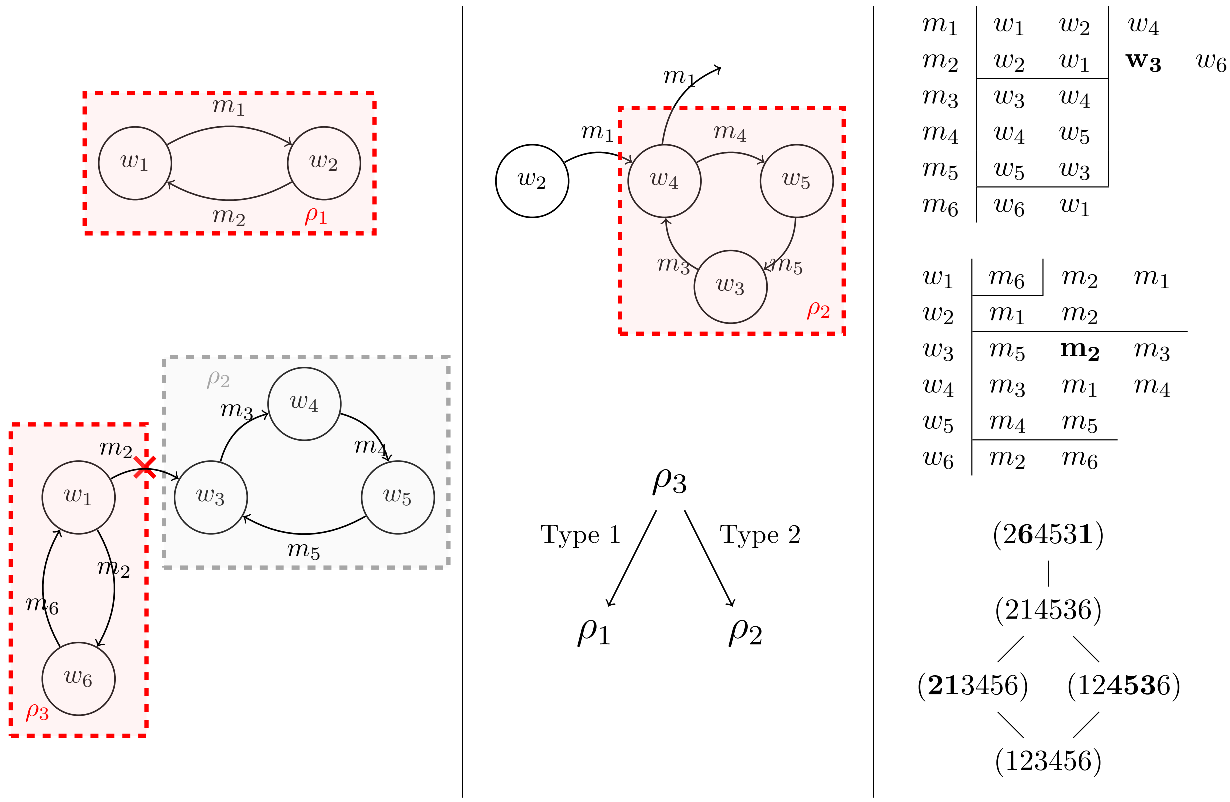



Figure 2: A sample instance with rotation poset and stable matching lattice.

Now, imagine running $MPDA(P_{w_1}(\mu_0))$ (where woman 1 truncate her list just above $m_1$). The rejection chain is $(w_1, m_1, w_2, m_2, w_1)$, and each woman recieves at most one proposals from a man she prefers to her match in $\mu_0$, so by claim 4.2 we've discovered the rotation $\rho_1 = [(w_1, m_1), (w_2, m_2)]$.

Next we run $MPDA(P_{w_2}(\mu_1))$. The rejection chain is $(w_2, m_1, w_4, m_4, w_5, m_5, w_3, m_3, w_4, m_1)$. As man $m_1$ failed to find a new partner, we know that the result cannot be stable (and indeed by claim 3.5, $m_1$ is the best stable partner of $w_2$). However, we still "learned something" along the way: $w_4$ accepted a proposal from both $m_1$ and $m_3$, and the rejection chain included her twice. If we had started the rejection chain from $w_4$, we could have actually gotten $MPDA(P_{w_4}(\mu_1)) = (214536)$, a new stable matching which differs from $\mu_1$ by the rotation $\rho_2 = [(w_4, m_4), (w_5, m_5), (w_3, m_3)]$. Thus, we can note this rotation $\rho_2$ and also the fact that $w_2$ has reached her best stable partner.

Because of the pair $(m_2, w_1)$, we know $\rho_1$ is a type 1 predecessor of $\rho_3$. Indeed, $\rho_1$ must be eliminated before $\rho_3$, because $w_1$ must be paired to $m_2$ before $\rho_3$ could possible happen[6].

Finally, the rejection chain of $MPDA(P_{w_1}(\mu_2))$ is simply $(w_1, m_2, w_6, m_6, w_1)$. This corresponds to the last rotation $\rho_3 = [(w_1, m_2), (w_6, m_6)]$. Because of the pair $(m_2, w_3)$, we know $\rho_2$ is a type 2 predecessor of $\rho_3$. Why does this mean that $\rho_2$ must be eliminated before $\rho_3$? Consider trying to eliminate $\rho_3$ before $\rho_2$, for example, by running $MPDA(P_{w_1}(\mu_1))$. The rejection chain is $(w_1, m_2, w_3, m_3, w_4, m_4, w_5, m_5, w_3, m_2, w_6, m_6, w_1)$, and $w_3$ got two proposals and accepted them both. The rejection chain between those two proposals corresponds to $\rho_2$. Thus, trying to eliminate $\rho_3$ *triggered* the elimination of $\rho_2$, even though none of the agents in $\rho_3$ appear in $\rho_2$.

[6] For an example of what would happen if you try to eliminate a type 1 predecessor before its successor, see figure 1.

### 5.2 Algorithm description

Given what we know, the high-level interpretation of algorithm 2 is fairly intuitive. The algorithm starts from the man-optimal stable outcome $\mu_0$. Along the way, it maintains a matching $\tilde{\mu}$, which is always stable, and initially set to $\mu_0$. From that point on, our goal is to make the smallest possible changes upward in the lattice $\mathcal{L}$, i.e. to make it so that each new value of $\tilde{\mu}$ covers the old value.

The algorithm works by picking any woman $\hat{w}$, and simulating the proposals and rejections made in $MPDA(P_{\hat{w}}(\tilde{\mu}))$ by having $\hat{w}$ "divorce" her husband and continuing to run deferred acceptance. By claim 3.6, we get a new matching which covers $\tilde{\mu}$ (i.e. we find a rotation) if and only if no woman receives multiple proposals from a man she prefers to her match in $\tilde{\mu}$. We cannot efficiently guarantee that this will hold for the $\hat{w}$ that we pick. However, we can get around this issue by using the following trick: when a woman considers a new proposal, she decides whether to accept as if she were still matched to her partner in $\tilde{\mu}$, even if she has already accepted a proposal that puts her above that man[7]. Then, whenever a woman $w^*$ receives a second proposal from a man she prefers to her match in $\tilde{\mu}$, we pause for a minute. We consider stable matching corresponding to the execution of $MPDA$ *between these two proposals which* $w^*$ *receives.* This corresponds to running $MPDA(P_{w^*}(\tilde{\mu}))$ and getting a matching $\mu'$ which covers $\tilde{\mu}$. So we reset $\tilde{\mu}$ to the equivalent of $\mu'$ and record the corresponding rotation in the graph $G$.

To prevent the algorithm from doing unnecessary repeated work, and to efficiently keep track of when we reach the woman-optimal outcome, we maintain a set $S$. Whenever a rejection chain starting with $\hat{w}$ ends in an unmatched woman or unmatched man, we know by claim 3.5 that $\hat{w}$ cannot receive a better stable match. Because we eliminate all cycles along the way, every woman after $\hat{w}$ on the rejection chain would also trigger this same event. Thus, each woman on the current rejection chain has reached their optimal match, and can be added to $S$. The algorithm runs until $S$ is all of $\mathcal{W}$.

Along the way, we keep track of type 1 and type 2 predecessor using a straightforward application of their definition. For the type 1 predecessors, it suffices to look at which rotations move the men, and create predecessor relations between each successive rotations moving the same man. For the type 2 predecessors, intuitively we detect under which conditions eliminating $\rho_2$ would *force* the elimination of $\rho_1$, because some woman who appears in $\rho_1$ would have accepted a proposal that a man $m$ makes as he moves through $\rho_2$. To implement this, we label the men on each woman's preference list, putting a label $\rho$ for each woman $w$ in $\rho$ and each man $m$ such that $\rho$ moves $w$ from below $m$ to above $m$. Then, we accumulate the corresponding rotations as the men make proposals (i.e. as a man $m$ gets rejected by some woman $w$, we label $m$ with any rotation that moved $w$ from below to above $m$).

### 5.3 Proof of correctness

Our main procedure is given in full detail as algorithm 2. An execution sequence of algorithm 2 is defined by the choice of rejections that the algorithm triggers, more specifically, by each choices of the woman $\hat{w}$ every time we reach line 7. As in the case of $MPDA$, we will see that the final result is independent of these choices and that the total amount of work done is $O(n^2)$.

[7] Again, figure 1 provides an example of why this is necessary. There, if $w_2$ compared $m_3$ against her match in $\mu$ (namely $m_1$) instead of in $\tilde{\mu}$ (namely $m_2$) then we would not find two distinct rotations $\rho_1$ and $\rho_2$. Instead, we would find the list $[(w_1, m_1), (w_2, m_2), (w_3, m_3)]$, which is not a rotation because eliminating it from $\mu_0$ does not result in a matching which covers $\mu_0$, i.e. we would *miss* some stable matchings in between $\mu_0$ and the next matching found in $\tilde{\mu}$.

**Algorithm 2** Finding the Rotations and Predecessor Graph

**Input** A stable matching instance with men $\mathcal{M}$ and women $\mathcal{W}$

**Output** A direct graph $G$ on the rotations of the instance

Let $\tilde{\mu}$ be the man-optimal stable matching from $MPDA$; Let $\mu$ be a copy of $\tilde{\mu}$
For each man $m$, let $R(m)$ be the set of women who rejected $m$ during the run of $MPDA$.
Let $S$ be the set of unmatched women in $\mu$
Set $pred_m^1 = \emptyset$ for each man $m$ ▷ Store the most recent rotation moving $m$
For each woman $w$ and each man $m$ on $w$'s list, label $m$ in the list with $\emptyset$
▷ Store rotations moving $w$ from *below* to *above* men
**while** $S \neq \mathcal{W}$ **do**
  Pick any $\hat{w} \in \mathcal{W} \setminus S$ ▷ These selections define an execution sequence
  Let $m = \mu(\hat{w})$; let $V = [(\hat{w}, m)]$
  Set $\mu(\hat{w}) = \emptyset$ and add $\hat{w}$ to $R(m)$ ▷ $\hat{w}$ rejects $m$
  Let $pred_m^2 = \emptyset$ ▷ Keep track of predecessor rotations
  **while** $V \neq [\ ]$ **do**
    Let $w \leftarrow$ NextAcceptingWoman($m$)
    **if** $w = \emptyset$ or $w \in S$ **then** ▷ No stable matching exists rotating partners in $V$
      Restore $\mu \leftarrow \tilde{\mu}$
      Add all women in $V$ to $S$; Set $V = [\ ]$
    **else if** $w$ appears in $V$ **then** ▷ New rotation found
      **if** $w \neq \hat{w}$ **then**
        Swap $\mu(w) \leftrightarrow m$ ▷ $w$ does not reject $m$ yet (see Claim 5.2, item 2))
      BuildNewRotation($w$)
    **else** ▷ Continue building rejection chain $V$
      Append $(w, \mu(w))$ to the end of $V$
      Swap $\mu(w) \leftrightarrow m$; Add $w$ to $R(m)$ ▷ $w$ rejects $\mu(w)$
      Let $pred_m^2 = \emptyset$ ▷ Keep track of predecessor rotations

**function** NextAcceptingWoman($m$)
  Let $w$ be $m$'s most preferred woman not in $R(m)$ (or $\emptyset$)
  **while** $w \neq \emptyset$ and $\tilde{\mu}(w) >_w m$ **do** ▷ while $w$ has received a better *stable* match
    Add $w$ to $R(m)$ ▷ $w$ rejects $m$
    If $w$ labeled $m$ with $\rho$, add $\rho$ to $pred_m^2$
    ▷ If rotation $\rho$ moved $w$ above $m$, then $\rho$ must precede the current rotation
    Update $w$ to $m$'s top woman not in $R(m)$ (or set $w$ to $\emptyset$)
  Return $w$

**function** BuildNewRotation($w$)
  Suppose $V = [(w_1, m_1), (w_2, m_2), \ldots, (w_k, m_k)]$ with $w = w_\ell$ for some $\ell \leq k$
  Update $\tilde{\mu}(w_i) = \mu(w_i)$ for each $i = \ell, \ell + 1, \ldots, J$ ▷ Eliminate new rotation $\rho^*$
  Remove $\rho^* = [(w_\ell, m_\ell), \ldots, (w_k, m_k)]$ from $V$
  Add rotation $\rho^*$ with type 1 predecessors $\bigcup_{i=\ell}^{k} pred_{m_i}^1$
  and type 2 predecessors $\bigcup_{i=\ell}^{k} pred_{m_i}^2$ to $G$
  **for** each $i = \ell, \ldots, k$ **do** Set $pred_{m_i}^1 = \rho^*$
  **for** each $i = \ell, \ldots, k$, and for each man
    $m$ between $m_i$ and $m_{i-1}$ (or $m_\ell$ and $m_k$ if $i = \ell$) on $w_i$'s list **do**:
    $w_i$ labels $m$ with $\rho^*$

**Claim 5.1.** *At any step of algorithm 2, every woman in the set $S$ has reached her optimal stable match.*

*Proof.* We prove this claim by induction on the number of iterations (of the outer loop in Algorithm 2 from line 7 to line 26) the algorithm has run. Let $V_i, \tilde{\mu}_i, S_i$ denote the value of $V, \tilde{\mu}, S$ at the end of iteration $i$ respectively.

Firstly, $S_0$ is the set of all unmatched women in MOSM. All women in $S_0$ has already reached their optimal match by the rural hospital theorem 2.4.

Next, assume for $k \le i$, every woman in $S_k$ has reached her optimal stable match at the end of iteration $k$. If $S_{i+1} = S_i$, then the same must hold for $S_{i+1}$. If $S_{i+1} \neq S_i$, then at the end of iteration $i+1$ the algorithm must have entered the if branch on line 13, which adds all women in $V$ to $S$ in line 15. Let $w(V_i)$ be the set of all woman in $V_i$. Then $S_{i+1} = S_i \cup w(V_{i+1})$. We claim that every woman in $w(V_{i+1})$ must have reached their optimal stable matching in $\tilde{\mu}_{i+1}$.

Observe that at any point where a woman $w$ receives two proposals from men she prefers to her match in $\tilde{\mu}$, the algorithm enters the if branch on line 20, where subroutine BuildNewRotation updates $\tilde{\mu}$. After the update, all women in $V$ have received at most one proposal from men she prefers to her match in $\tilde{\mu}$. Thus only the last woman in the rejection chain could have received more than one proposal from men she prefers to her in $\tilde{\mu}_{i+1}$. However, the last woman is either $\emptyset$ or in $S_i$, and is never added to $V_{i+1}$. Therefore every woman in $w(V_{i+1})$ has only received one proposal from men she prefers to her match in $\tilde{\mu}_{i+1}$.

Let $V_{i+1} = [(w_1, m_1), (w_2, m_2)...(w_t, m_t)]$. Since $V$ represent the rejection chain in $MPDA(P_{w_1}(\tilde{\mu}_{i+1}))$, we know that $MPDA(P_{w_1}(\tilde{\mu}_{i+1}))$ results in an unstable matching. Moreover, since all woman $w_j \in w(V_{i+1})$ only receives one proposal (from men $m_{j-1}$) that she prefers to her match in $\tilde{\mu}_{i+1}$, $MPDA(P_{w_j}(\tilde{\mu}_{i+1}))$ must have the rejection chain $[(w_j, m_j), ...(w_t, m_t)]$, and also result in an unstable matching. By Claim 3.5, $w_j$ is pair with her optimal stable partner in $\tilde{\mu}_{i+1}$ already. □

**Claim 5.2.** *Algorithm 2 terminates and runs in $O(n^2)$ time.*

*Proof.* Thoughout the algorithm, at each time step, one of the following events happen: (1) a woman rejects a man $m$, (2) a man propose to a woman $w$, either after being rejected or the man repropose after a rotation has been built, (3) A new rotation $\rho$ is extracted, (4) an earlier rotation is added as type 1 or type 2 predecessor of new rotation $\rho$ and (5) women are added to $S$. Each event above takes constant time. We show below that the total number of above events is $O(n^2)$.

1) Woman $w$ can only reject man $m$ once. Thus in total event (1) happen $O(n^2)$ times.
2) We'd like to say that (as in the case of deferred acceptance) every man proposes to every woman at most once. However, there is an important exception to this: when a man $m$ is tentatively matched to a woman in $\mu$ (i.e. $m$ is in $V$), but the woman receives a new proposal which she accepts, then the man must propose to the woman again after the corresponding rotation has been created. However, this can happen at most once for every rotation, and by claim 4.3, there are $O(n^2)$ rotations. So the total number of proposals made is still $O(n^2)$.
3) There are $O(n^2)$ rotations, and each rotation is found at most once.
4) Throughout the algorithm, a rotation $\rho$ is added to $pred_m^1$ only when $\rho$ changes $m$'s stable partner to some woman $w$. $m$ can only have $O(n)$ different stable partners, and there are only $O(n)$ men. Thus a rotation can only be added as a type 1 predecessor $O(n^2)$ times. Similarly, we can count the number of type 2 edges. A rotation $\rho$ is added to $pred_m^2$ only when $w$ labeled $m$ with $\rho$, and $w$ rejects $m$. Moreover, each man on $w$'s list receives at most one label. So a rotation is added as a type 2 predecessor $O(n^2)$ times.

5) A fixed woman $m$ can only be added to $S$ once. Thus event (5) occur $O(n)$ times.

As we do a constant amount of work for all these events, we conclude that the execution time is $O(n^2)$.

□

We now prove that algorithm 2 traverses a maximal chain from the man-optimal to the woman-optimal stable outcome. Using the theory built up in section 4, this will allow us to fairly easily prove that algorithm 2 correctly outputs all rotations $\Pi$ and the predecessor digraph $G(\Pi)$.

**Claim 5.3.** *During the execution of algorithm 2, let the* $MPDA(P) = \mu_0, \mu_1, \ldots, \mu_k$ *denote the values* $\tilde{\mu}$ *in order, and let* $\rho_0, \rho_1, \ldots, \rho_{k-1}$ *denote the values of* $\rho^*$ *inserted into* $G$ *in order. Then each* $\mu_i$ *is a stable matching,* $\mu_k$ *is the woman-optimal stable match, and* $\mu_{i+1}$ *is the elimination of* $\rho_i$ *from* $\mu_i$ *for each* $i$*. In other words,* $\mu_0, \mu_1, \ldots, \mu_n$ *is a maximal chain in the stable matching lattice* $\mathcal{L}$*, with corresponding rotation sequence* $\rho_0, \rho_1, \ldots, \rho_{k-1}$*.*

*Proof.* First, we show by induction that each $\mu_i$ is stable. Assume that $\tilde{\mu}$ is stable and the BUILD-NEWROTATION function is called on line 19. Let $w^*$ denote the current value of $w$ at that line (so $w^*$ is the woman in $V$ who just accepted a proposal from a man she prefers to her match in $\tilde{\mu}$), and let $\rho^*$ be as on line 34. Consider the sequence of rejections made after the last time $\tilde{\mu}$ was changed. No woman in $\rho^*$ other than (possibly) $w^*$ has received multiple proposals from a man she preferred to her match in $\tilde{\mu}$. Furthermore, the sequence of rejections and proposals made between the first occurrence of $w^*$ in $V$ is exactly those made in $MPDA(P_{w^*}(\tilde{\mu}))$, i.e. the rejection chain of $w^*$ starting from $\tilde{\mu}$ corresponds exactly to $\rho^*$. By 3.6, $\mu' = MPDA(P_{w^*}(\tilde{\mu}))$ covers $\tilde{\mu}$. But the new value of $\tilde{\mu}$, set on line 33, is exactly $\mu'$ (which is exactly the elimination of rotation $\rho^*$ from $\tilde{\mu}$). Thus, $\mu_{i+1}$ covers $\mu_i$ for each $i$, and $\mu_{i+1}$ is the elimination of $\rho_i$ from $\mu_i$.

Algorithm 2 terminates only when $S$ is all of $\mathcal{W}$. But by claim 5.1, $S$ consist only of women who have reached their optimal stable match in $\tilde{\mu}$. Thus, when the algorithm terminates, $\tilde{\mu}$ is the woman-optimal stable match.

□

**Claim 5.4.** *Every rotation of* $\Pi$ *is found and put into* $G$ *over the course of algorithm 2. Furthermore, if* $\rho_1$ *is a predecessor of* $\rho_2$ *(type 1 or type 2) then* $\rho_1$ *will be found before* $\rho_2$*. Moreover, the set of predecessors in* $G$ *of every rotation* $\rho$ *are exactly the type 1 and type 2 predecessors defined above, i.e.* $G = G(\Pi)$*.*

*Proof.* The first two statements now readily follow from the previous claim and claim 4.5 (parts 2 and 3 respectively).

We know that for each rotation $\rho^*$, each predecessor of $\rho^*$ has certainly been found by the time we construct $\rho^*$. We now show that the predecessors of $\rho^*$ are appropriately marked. The type 1 predecessors are added on line 35, and they are exactly the most recently found rotations moving man $m$ (as consistently updated on line 37). But the most recent rotation moving $m$ must be the unique rotation which moved $m$ to his current match $w$, where $(w, m)$ appears in $\rho^*$. So the type 1 predecessors of $\rho^*$ are accurately marked.

For the type 2 predecessors, consider a man $m$ who appears in $\rho^*$ and moves from $w$ to $w'$. From the time $m$ entered $V$ (in line 10 or 23), we added to $pred_m^2$ (on line 28) each rotation $\rho$ such that $m$ was rejected by a woman $w$ and $\rho$ moved $w$ from below $m$ to above $m$ (all such rotations are predecessors of $\rho^*$, and thus have already been found and appropriately marked on lines 39 to 40). As $m$ is rejected by each woman between $w$ and $w'$ on his list, this covers all possible type 2 predecessors of $\rho^*$, so the type 2 predecessors are accurately marked on line 36.

□

We can now immediately conclude from claims 5.2 and 5.4 that the rotation predecessor graph $G(\Pi)$ can be correctly and efficiently computed.

**Theorem 5.5.** *Algorithm 2 computes $G(\Pi)$ in $O(n^2)$ time.*

## References


[AG18] Itai Ashlagi and Yannai A Gonczarowski. Stable matching mechanisms are not obviously strategy-proof. *Journal of Economic Theory*, 177:405–425, 2018.

[AKL17] Itai Ashlagi, Yash Kanoria, and Jacob D. Leshno. Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1):69 – 98, 2017.

[GI89] Dan Gusfield and Robert Irving. *The stable marriage problem: Structure and algorithms*. The MIT Press, Cambridge, MA, 1989.

[GNOR19] Yannai A Gonczarowski, Noam Nisan, Rafail Ostrovsky, and Will Rosenbaum. A stable marriage requires communication. *Games and Economic Behavior*, 2019.

[Gon14] Yannai A. Gonczarowski. Manipulation of stable matchings using minimal blacklists. In *ACM Conference on Economics and Computation, EC '14, Stanford , CA, USA, June 8-12, 2014*, page 449, 2014.

[IL86] Robert W Irving and Paul Leather. The complexity of counting stable marriages. *SIAM Journal on Computing*, 15(3):655–667, 1986.

[ILG87] Robert W Irving, Paul Leather, and Dan Gusfield. An efficient algorithm for the optimal stable marriage. *Journal of the ACM (JACM)*, 34(3):532–543, 1987.

[IM05] Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 53–62, 2005.

[Irv85] Robert W Irving. An efficient algorithm for the stable roommates problem. *Journal of Algorithms*, 6(4):577–595, 1985.

[KGW18] Anna R. Karlin, Shayan Oveis Gharan, and Robbie Weber. A simply exponential upper bound on the maximum number of stable matchings. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 920–925, New York, NY, USA, 2018. ACM.

[Pit19] Boris Pittel. On likely solutions of the stable matching problem with unequal numbers of men and women. *Mathematics of Operations Research*, 44(1):122–146, 2019.

[Rot86] Alvin E. Roth. On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica*, 54(2):425–427, 1986.

# A Contrasting Our Proofs With Those of [GI89]

The primary difference between our approach and the original is that [GI89] starts by defining a partial order on the set of rotations $\Pi$, then constructs a directed acyclic graph $G$ and proves that the transitive closure of $G$ gives the correct order on $\Pi$. In contrast, we start with $G = G(\Pi)$ and are able to prove directly that $G(\Pi)$ represents the set of all stable matchings.

Our definition of rotations is similar to that of [GI89]. The only difference is that we define rotations to be the difference between two stable matchings where one covers the other, whereas [GI89] essentially defines rotations as rejection chains of MPDA (where no woman receives multiple proposals from a better man than her old match) without explicitly mentioning MPDA (GI89 must prove that rotations give covering relations in their lemma 2.5.5, although they do not explicitly use the term "covering").

[GI89] defines a partial order relation on $\Pi$ via the partial order relation on their "minimal difference on a ring of sets". In turn, the ordering on the minimal differences is defined via the lattice ordering, restricted to the "irreducible elements (other than the man-optimal)". Building the theory of these partial orders occupies [GI89] for the entirety of their Chapter 2 (pages 67 to 102). We believe that intuition is lost through these layers of definition.

There are two steps to proving that the transitive closure of $G$ is exactly the partial order on $\Pi$. First, every edge in $G$ should be related in $\Pi$. We find that the essence of the proof in [GI89] (lemma 3.2.3) goes through without referencing a partial order on $\Pi$ at all. We capture this with our claim 4.5, which shows that any "valid" sequence of rotations (i.e. one which corresponds to a maximal chain) respects $G$.

The second step is to show that every relation in $\Pi$ is in the transitive closure of $G$. [GI89] heavily relies on the existing partial order on $\Pi$ for this step of their proof (lemma 3.2.4) because they prove that any *immediate* predecessor (i.e. a covering relation) in $\Pi$ must be related in $G$.

We are able to skip this crucial reliance on an order on $\Pi$ via our claim 4.7, which proves that every ordering of rotations which respects $G$ corresponds to a maximal chain in the stable matching lattice. The key lemma we use is claim 4.6, which shows that if two rotations are adjacent in some topological sort and do not have an edge between them, then that pair of rotations can be swapped (while maintaining the property that the topological sort corresponds to a maximal chain).

In the end, our method amounts to showing that the collection of topological sorts of $G$ is exactly the collection of linear extensions of $\Pi$. The "linear extensions are topological sorts" direction is equivalent to proving that all edges in $G$ are related in $\Pi$. The "topological sorts are linear extensions" direction is proved using *only* the "swapping lemma" (claim 4.6). The linear extensions of $\Pi$ are particularly natural objects in our case (as they correspond to maximal chains in the stable matching lattice). To the best of our knowledge, this "swapping lemma" strategy for proving that a graph gives a certain transitive closure has not been used before. It may be useful in other situations involving distributive lattices where maximal chains in the lattice are easy to reason about.

As in [GI89], we deliberately avoid mentioning Birkhoff's representation theorem. While this classical theorem immediately shows the existence of *some* partial order which represents any distributive lattice, it does not show how to find this representation or give any structure regarding what the elements of the partial order are. Indeed, Birkhoff's theorem by itself could not even show that the partial order $\Pi$ is polynomial-size.

# B A Minor Error in [GI89]

The *minimal-differences* algorithm of [GI89] (more precisely, figure 3.2 on page 110) correctly identifies all of the rotations in a stable matching instance. However, there is a slight error in the construction of the order relations for the rotation poset. In particular, once the rotations are found (via an algorithm essentially equivalent to our algorithm 2, but without keeping track of predecessor relations), they propose Algorithm 3 as shown below.

**Algorithm 3** Construct predecessor relations

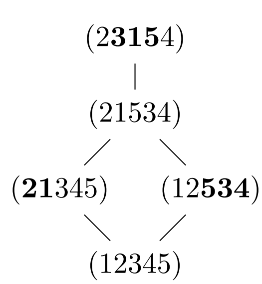
**for** Each rotation $\rho$ and pair $(m_i, w_i) \in \rho$ **do**
  Label $w_i$ in $m_i$'s preference list with a type 1 $\rho$ label
  **for** Each $m$ strictly between $m_i$ and $m_{i-1}$ on $w_i$'s list **do**
    Label $w_i$ in $m$'s preference list with a type 2 $\rho$ label
**for** Each man $m$ **do**
  Set $\rho^* = \emptyset$
  **for** Each woman $w$ on $m$'s preference list, in order **do**
    **if** $w$ has a type 1 label of $\rho$ **then**
      **if** $\rho^* \neq \emptyset$ **then** Add $\rho^*$ as a predecessor of $\rho$
      Set $\rho^* = \rho$
    **if** $w$ has a type 2 label of $\rho$ **then**
      **if** $\rho^* \neq \emptyset$ **then** Add $\rho$ as a predecessor of $\rho^*$

The idea behind this algorithm is reasonable: certainly the type 1 labels in any man's chain should be related in the poset (as the man needs to reach a certain partner before the next rotation can be found). Furthermore, suppose a type 2 label $\rho_2$ is between two type 1 labels, $\rho_1$ and $\rho_3$. We know that $\rho_1$ moved $m$ from his partner in $\rho_1$ to his partner is $\rho_3$, as men propose in their preference order at most once to each woman. Along the rejection chain from his partner in $\rho_1$ to his partner in $\rho_3$, $m$ would propose to some woman $w$ in $\rho_2$, and $w$ likes $m$ better than her match in $\rho_2$. Thus, the rejection chain of $\rho_1$ will certainly trigger $\rho_2$, and $\rho_1$ must be a predecessor of $\rho_2$.

However, the above reasoning fails in certain cases. Namely, in the case where there is a type 1 label $\rho^*$ followed by a type 2 label $\rho$ on woman $w$, but in $\rho^*$ man $m$ does not move from above $w$ to below $w$. In this case, the rejection sequence $\rho^*$ does not actually trigger rotation $\rho$.

| | | | |
|---|---|---|---|
| $m_1$ | 1 | 2 | **3** |
| $m_2$ | 2 | 1 | 3 |
| $m_3$ | 3 | 5 | 1 |
| $m_4$ | 4 | 3 | 5 |
| $m_5$ | 5 | 4 | |

| | | | | |
|---|---|---|---|---|
| $w_1$ | | 3 | 2 | 1 |
| $w_2$ | | | 1 | 2 |
| $w_3$ | 2 | 4 | **1** | 3 |
| $w_4$ | | | 5 | 4 |
| $w_5$ | | 4 | 3 | 5 |




| | | | |
|---|---|---|---|
| $m_1$ | 1 | 2 | 3 |
| | $T1\rho_1$ | | $T2\rho_2$ |
| $m_2$ | 2 | 1 | 3 |
| | $T1\rho_1$ | $T1\rho_3$ | |
| $m_3$ | 3 | 5 | 1 |
| | $T1\rho_2$ | $T1\rho_3$ | |
| $m_4$ | 4 | 3 | 5 |
| | $T1\rho_2$ | $T1\rho_3$ | |
| $m_5$ | 5 | 4 | |
| | $T1\rho_2$ | | |

Figure 3: A tricky case for algorithm 3, including the labeled preference lists.

For a concrete counterexample, consider the stable matching instance in figure 3, drawn along-side its lattice $\mathcal{L}$ (with matchings written as the by writing the partner of $w_1, w_2, \ldots, w_5$ in order). The rotations of this instance are $\rho_1 = [(1,2),(2,1)]$, $\rho_2 = [(3,3),(4,4),(5,5)]$, and $\rho_3 = [(2,1),(4,3),(3,5)]$, and $\rho_1$ and $\rho_2$ are both type 1 predecessors of $\rho_3$. The labeled preference list of the men, given by applying algorithm 3, is also drawn in figure 3.

The above algorithm causes $\rho_1$ to be marked as a predecessor of $\rho_2$, even though they are independent and both exposed in the man-optimal stable matching. Our algorithm 2 circumvents this problem by storing the type 1 and type 2 labels in different places, and detecting the required orderings on the rotations more directly.

Another way around this problem, which is more similar to [GI89]'s algorithm 3, would be to write "type 1 end markers" for the final type 1 label in each man's preference list (note that this problem can only happen for type 2 labels after the final type 1 label, because if there is another type 1 label after the type 2 label, the man must actually move below the woman $w$ where the type 2 label was marked). More specifically, for the last type 1 label on $m$'s list, say of a rotation $\rho$, mark "type 1 end" on the woman $w$ for which $\rho$ moves $m$ to $w$. Then, ignore any type 2 labels after the "type 1 end" mark.