Efficiently Representing the Set of Stable Matchings by the Rotation Poset

Clay Thomas
claytont@princeton.edu

August 10, 2019

Abstract

Stable matchings are a classical area of study, and the well-known treatment of [Gusfield and Irving. "The stable marriage problem: structure and algorithms". MIT press, 1989] gives strong structural results useful for understanding this problem. Namely, the set of stable matchings for an $n \times n$ instance is in a bijection with the set of downward-closed subsets of a certain polynomial-sized partial order (the elements of which correspond to the set of stability-preserving "partner rotations"). While Gusfield and Irving's treatment is quite elegant, the proofs given are a bit long and abstract. In this paper, we give a slight modification of the Gusfield and Irving algorithm, emphasizing its conceptual basis in the same principles as differed acceptance. We prove correctness mostly by using concrete algorithmic properties. We also provide a more streamlined and unified construction of the partial ordering relations, correcting some slight errors in the original text. Our algorithm naturally handles unacceptable pairs and unequal numbers of men and women.

We also discuss two recent advancements in the theory of *counting* stable matchings. First, using basic properties of the partial order on rotations (which we reprove), a exponential worst-case upper bound on the number of stable matchings has been shown in [Karlin, Gharan, and Weber. "A simply exponential upper bound on the maximum number of stable matchings." STOC, 2018]. Second, upper bounds for the *expected* number of stable matchings when there are an unequal number of men and women have recently been found in [Ashlagi, Kanoria, and Leshno. "Unbalanced random matching markets: The stark effect of competition." Journal of Political Economy, 2017] (it turns out there are far fewer matches than in the balanced case). For the second result, we provide greatly simplified proofs of some preliminary results in this direction.

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 matches are made between members of different groups. The most commonly used stable matching mechanism is "one-side proposing differed acceptance", which has the nice properties of being simple to implement, fast to execute, and strategyproof for the proposing side [DF81] (moreover, it is the unique mechanism satisfying strategyproofness for one of the sides [GS85]).

Many structural properties are known about the set of stable matches for a given instance (known as the *core* of the instance). For instance, the core forms a distributive lattice with a compact (and algorithmically friendly) representation, and every distributive lattice is the core of some "not too large" stable marriage instance [IL86]. This latter fact means that, a priori, one cannot say much more about the core than one can say about arbitrary distributive lattices[1]. However, in addition to the structure that the core must necessarily have, it is very interesting to know what qualities the core *probably* has.

The "first" randomized model of stable matching is to take preferences completely uniform for each agent. For a while, it's been known that in a randomized balanced market, i.e. one with the same number of agents on each side, there's probably a large core and a big imbalance between the different sides – the proposing side gets matched which they favor significantly more [Pit89]. However, in a recent breakthrough paper [AKL17], it was found that this distinction *almost vanishes* in large markets when there is an imbalance of just *one* more agents on one side than the other. In this note we survey this result.

The basic setup of stable matching has many variations. The central focus of this note is the setting with an unequal number of agents on each side. Some other import variations include adding "unacceptable matches" (agents do not rank every agent on the other side), "indifference" (agents sometimes don't distinguish their preference between some matches), "many-to-one matchings" (where several agents from one side are matched to one on the other), and "couples" (where certain agents on one side express their preferences/acceptable matches jointly). We use the "unacceptable matches" concept as a technical tool, but ignore the subtleties of the other variations.

Relation to prior work. No claim in this paper is original, and much of our treatment and proofs are very similar to that presented in [GI89]. The point where we start to differ is in proving results about the rotation poset: we use the concepts of women truncating their preference lists and running MPDA to prove most of our claims. This approach has been surveyed in [**?**].

==== HERM actually idk things were pretty clear to begin with in all these proof... but whatever. I'm still not sure how different my algorithm is from anything else.

Organization. In section 2, we review the basic properties of differed acceptance and stable matchings, which will motivate some of the techniques to come. In section 3, we review the case of a balanced market, and discuss which parts of the analysis fail for unbalance markets. In section 4, we prove some results on unbalanced markets, taking a new proof approach based around differed acceptance which is closer to the analysis for the balanced market. In section 5, we discuss the proof technique given in [AKL17].

[1] One recent result which runs counter to this statement is given in [KGW18], where the authors use the size of the stable marriage instance to bound the number of elements in the distributive lattice of the core. There, the exact meaning of the words "not too large" becomes important.

2 Review: Differed Acceptance and Stable Matching

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 $\emptyset \succ_m w$.

=== More defining to do here:: A matching is a one-to-one assignment of men to women, which we denote by $\mu : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W} \cup \{\emptyset\}$. We write $\mu(i) = \emptyset$ if agent i is unmatched.

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

The man-proposing differed acceptance algorithm is given by Algorithm 1. We provide simple proof of the basic properties of this algorithm.

Algorithm 1 MPDA: Men-proposing differed acceptance algorithm

```
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 μ. 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.

This next claim allow us to easily prove several corollaries. The proof follows this strategy: 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 μ be any matching. Suppose that some pair, matched in μ, 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 μ, then μ 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 μ, m' has not yet proposed to w'. Because men propose in their preference order, we have $w \succ_{m'} w'$. However, this means μ is not stable.

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

We can now formalize our intuition that DA 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. *In the match returned by MPDA,*

1. *every $m \in \mathcal{M}$ is paired to his most preferred stable partner.*
2. *every $w \in \mathcal{W}$ is paired to their worst stable match in $\mathcal{M}$.*

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

The matching output by the MPDA algorithm is independent of the order in which men are selected to propose.

Using our results so far, we can prove the following weaker version of the rural hospital theorem[2] which will be key for some of our later results.

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] properly refers to many-to-one matching markets are considered (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 always recieve the same doctors (and same number of doctors) in *every* stable outcome.

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

The final results of this section strengthens the intuition provided by claims 2.3 and 2.3 which states that the incentives of women and men are exactly opposed over the set of stable matchings.[3]

Claim 2.5. *Let μ, μ' 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 μ' is stable, yet m and w are not matched in μ', 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 μ' 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 μ, and in the second case, (m, w') form a blocking pair in μ. Thus, in either case μ is not stable. $\square$

Claim 2.6. *Let μ and μ' be stable matchings. Every man (weakly) prefers μ over μ' if and only if every woman (weakly) prefers μ' over μ.*

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. $\square$

3 The Lattices of Stable Matchings

A partial order $\leq$ is a reflexive, transitive, antisymmetric relation. 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 meet of a and b and $a \wedge b$ the join 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 lattice L is distributive if the join and meet operations satisfy the following equations:

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

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

An element a of a lattice covers an element b, denoted $a \gtrdot b$, when $a > b$ and no element c exists with $a > c > b$.

[3] DO NOT INCLUDE: note for myself: I think this fails in a subtle way for the resident-hospitals problem... look into this.

For stable matches μ, μ', we say that μ *dominates* μ' if, for every $m \in \mathcal{M}$, we have $\mu(m) \succeq_m \mu'(m)$, that is, if every man is at least as happy with his match in μ as in μ'. If μ dominates μ' we write $\mu \leq \mu'$. (ORGANIZE THIS THOUGHT OR POSSIBLY REVERSE IT: we (somewhat misogynistically) say that one matching dominates another based on the men's preferences. However, we visualize starting at the man-optimal stable outcome at the bottom of the stable matching lattice, and the woman-optimal being at the top. This is why we write $\mu \leq \mu'$ when μ dominates μ')

Theorem 3.1. *The collection $\mathcal{L}$ of all stable matchings of some instance form a distributive lattice under the dominance ordering $\leq$.*

Proof. It's easy to see that $\leq$ forms a partial order on $\mathcal{L}$. We'll show that least upper bounds exists (the proof for greatest lower bounds is identical, interchanging men with women). For stable matchings μ, μ', define $\tilde{\mu} = \mu \vee \mu'$ such that, for each woman w, $\mu(w)$ is the most preferred partner of w among $\mu(w)$ and $\mu'(w)$. It's clear that, if $\tilde{\mu}$ is a stable matching, then it is the least upper bound for μ and μ'.

First, we claim that $\tilde{\mu}$ is a matching. Suppose 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 μ or μ', 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 μ' is not stable.

Finally, we show that the join and meet operations in $\mathcal{L}$ are distributive. Analogously to the above, we would define $\mu \wedge \mu'$ such that every man gets their preferred partner from μ or μ'. By claim 2.5, this is equivalent to defining $\mu \wedge \mu'$ such that every woman gets worst partner from μ or μ'. Thus, the join and meet operations are distributive for the same reason that the operations of min and max distribute over each other. Thus, for every woman w,

$$(\mu_1 \wedge (\mu_2 \vee \mu_3))(w)$$

equal to the worse of $\mu_1(w)$ and (the better of $\mu_2(w)$ and $\mu_3(w)$). On the other hand, we have the better of (the worse of $\mu_1(w)$ and $\mu_2(w)$) and (the worse of $\mu_1(w)$ and $\mu_3(w)$), which with a little thought you can realize that these two equations are equal. □

We will not directly use this theorem, but it serves as motivation for why one might expect a compact representation of $\mathcal{L}$ to exist. In general, when one encounters a distributive lattice, it's always useful to ask what its join irreducible elements are, and see if there's a natural mathematical or algorithmic interpretation.

Theorem 3.2 (Birkhoff's representation theorem)**.** *For any finite distributive lattice L, there exists a subset P of L (namely, the "join irreducible elements" of L, which are those such that, whenever $a \vee b = c$, we have $a = c$ or $b = c$) such that L is isomorphic to the collection of downward-closed subsets of P (under the partial order induced by restricting L to P), which forms a lattice under set containment.*

Claim 3.3. *Let P be some instance of the stable matching problem with man-optimal stable match μ_0, and let P' be an instance identical to P, except some set of women truncate their preference lists. Let μ be the result of $MPDA(P')$. Then μ is stable for the original set of preferences P if and only if the set of matched agents in μ is identical to that in μ_0.*

Proof. One fact towards this: Observe that $MPDA$ produces the man optimal stable match such that each woman (if matched) receives a match higher than her truncation point. □

Claim 3.4. *Let μ be a stable matching for preferences P. For each woman w, let P_w denote the preference profile where each preference list is identical except for that of w, and w truncates her list from P one place after her stable partner from μ. The rest of the women should truncate their lists just BEFORE the match in μ. Then the collection of matches which cover μ is contained in the set $\{\mu' | \mu' = MPDA(P_w), \mu' \in \mathcal{L}\}_{w \in \mathcal{W}}$, and all of those elements dominate μ'.*

Proof. Key observation for the next section: the sequence of rejections made (ignoring terminal phases) are a valid sequence of rejections for the original MPDA algorithm with women truncating their preferences. So you can find a maximal path from bottom to top of the lattice. □

Claim 3.5. *If $\mu' \gtrdot \mu$ is a covering pair in the stable matching lattice then μ' and μ differ by a rotation.*

Proof. In terms of what we need to eventually show, this won't be quite enough. E.g. $(12)(23) = (132)$ is a simple rotation when considered as a permutation, but its composed of two (comparable) flips and not a simple rotation in the stable marriage instance.

Here's something a bit more like it: bring this analysis after the algorithm, then say this: "if V is a simple cycle, then μ_{new} covers μ_{old}, because (clearly its $\geq$) and in order to be $> \mu_{old}$, some woman must reject her match.

As $\mu' > \mu$, some woman w must recieve a strictly better match in μ' than in μ. □

Q: why must the set of rotations found be the same for every execution of the algorithm? One approach to proving this formally: consider execution sequences P and Q which contain different rotations, say $\rho \in P \setminus Q$.

4 The Rotation Poset

The algorithm starts from the man-optimal stable outcome (MOSM) by running differed acceptance. Then it triggers a series of "divorces", i.e. it breaks a marriage $(m, \hat{w})$ and "continues running differed acceptance". This involves m again proposing down his list until accepted by some woman who prefers m to her current best known stable match. This "rejection chain" can involve cycles, where a woman accepts a proposal for the second time. Ignoring cycles for now, the chain can end in one of two ways: a "terminal phase" in which an element of $\overline{\mathcal{W}}$ (the set of women unmatched in the MOSM) receives a match, or an "improvement phase" in which $\hat{w}$ receives a better stable match. A terminal phase clearly cannot constitute a stable match, by the rural hospital theorem (claim 2.4). On the other hand, [AKL17] shows that any improvement phase yields a new stable outcome.

Some notes:

Algorithm 2 MOSM to WOSM Conversion Algorithm

Let μ be the man-optimal stable matching
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 $\overline{\mathcal{W}}$
while $S \neq \mathcal{W}$ **do**
  Store $\tilde{\mu} \leftarrow \mu$ ▷ $\tilde{\mu}$ stores the most recent *stable* match encountered
  Pick any $\hat{w} \in \mathcal{W} \setminus S$
  Let $m = \mu(\hat{w})$; let $V = [(m, \hat{w})]$
  Set $\mu(\hat{w}) = \emptyset$ and add $\hat{w}$ to $R(m)$ ▷ $\hat{w}$ rejects m
  while $V \neq [\,]$ **do**
    Let $pred_m = \emptyset$ ▷ We keep track of the rotations
    Let w be m's most preferred woman not in $R(m)$
    while $\tilde{\mu}(w) >_w m$ **do** ▷ while w has received a better stable match
      Add w to $R(m)$ ▷ w rejects m
      If rotation ρ moved w above m, add ρ to $pred_m$
      Update w to m's top woman not in $R(m)$ (or set w to $\emptyset$)
    (So now $m >_w \tilde{\mu}(w)$ (if $w \neq \emptyset$))
    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 $w \neq \hat{w} = v_1$ **then**
        Set $m_{float} \leftarrow \mu(w)$
      Suppose $V = [(m_1, w_1), (m_2, w_2), \ldots, (m_k, w_k)]$ with $w = w_\ell$ for some $\ell \leq k$
      Update $\mu(w) = m$ and $\tilde{\mu}(v_i) = \mu(v_i)$ for $i = \ell, \ell + 1, \ldots, J$
      Remove $\rho = [(m_\ell, w_\ell), \ldots, (m_\ell, w_\ell)]$ from V
    else if $w \notin V$, $w \notin S$ **then** ▷ Continue building rejection chain V
      Swap $\mu(w) \leftrightarrow m$; Add w to $R(m)$ ▷ w rejects $\mu(w)$
      Append (m, w) to the end of V

- The most recent stable match $\tilde{\mu}$ is stored, while the "partial match" μ may not correspond to any stable match.

- When $w \in V$, we compare m against w's match in $\tilde{\mu}$, not in μ. This is because we want to make the smallest changes possible that result in new stable outcomes, and w will still make an improvement from the old stable outcome by accepting this m.

- If a cycle occurs during a rejection chain, that cycle is immediately "recorded" in $\tilde{\mu}$. This is because this is equivalent to picking $\hat{w}$ to be any woman in the simple cycle, and (similar to differed acceptance) the execution of this algorithm is independent of the order in which women are picked.

- We actually don't need to wait until the rejection chain hits $\overline{\mathcal{W}}$. Instead, we let S denote the set of women for whom we know that a phase starting with $\hat{w} \in S$ will be terminal. If at any point a woman in S then accepts the proposal, we know that the phase we are currently in will be terminal as well.

Formally, [AKL17] proves that the following all hold during the run of algorithm 2:

- $\tilde{\mu}$ is always a stable matching.

- When a phase is terminal, $\hat{w}$ is currently matched in $\tilde{\mu}$ to her optimal stable partner. Because we eliminate cycles, any other woman in V would also cause a terminal phase if she was chosen as $\hat{w}$. Thus, S always consists of women who have reached their optimal stable match.

- Every woman is at some point in time matched to every stable partner she has, in order from her least preferred to her most preferred.

- Algorithm 2 terminates with $\tilde{\mu}$ being the woman-optimal stable outcome.

5 Balanced Random Markets

For a man m and woman w in some preference set P, define $R_m(w) := |\{w' : w' \succeq_m w\}|$, i.e. the number of women preferred at least as much as w, i.e. the index of w on m's ordered preference list. We will say that m ranks w better than w' when $R_m(w) < R_m(w')$. Given a matching μ stable under preferences P, define

$$R_{men}(\mu) := \frac{1}{|\mathcal{M} \setminus \overline{\mathcal{M}}|} \sum_{m \in \mathcal{M} \setminus \overline{\mathcal{M}}} R_m(\mu(m))$$

That is, $R_{men}(\mu)$ is the average over matched men of the men's rank of their wives. Define $R_w(m)$ and $R_{women}(\mu)$ analogously.

A random matching market is defined by a set of randomly draw preferences $P = P_{n,m}$ with n men and m women, where each of the n men with uniformly random rankings over the m women, and m women with uniformly random rankings over the n men. Given a set of preferences P, let $MOSM(P)$ denote the man-optimal stable outcome, i.e. the result of running MPDA (men-proposing differed acceptance). Likewise, let $WOSM(P)$ denote the woman-optimal stable outcome, i.e. the result of running women-proposing differed acceptance.

The crucial observations to dealing with MPDA are the following:

- The sum of the ranks that the men have for their wives in the MOSM is exactly the number of proposals made during MPDA.
- When $n \leq m$, MPDA terminates as soon as n distinct women are proposed to. Thus, the sum of ranks of the men is essentially a *coupon collector* random variable, with expectation $O(\log n)$.

Proposition 5.1. *Let P be a random matching market with n men and $n+k$ women for $k \geq 0$. Then $\mathbb{E}\left[R_{men}(MOSM(P))\right] = O(\left(\frac{n+k}{n}\right) \log\left(\frac{n+k}{k}\right))$.*

Proof. Consider running men-proposing differed acceptance on M. At the steps requiring some unmatched man to be selected, select one uniformly at random. Let Y denote the random variable giving the total number of proposals made by men to women. By the so-called "principle of differed decisions", running this algorithm on a random matching market is equivalent to running the following randomized algorithm: women's preferences are taken as input, but men's are not, and whenever a man is ready to propose to a woman, simply select a woman he hasn't yet proposed to uniformly at random.

Let Z denote a "coupon collector" random variable defined by the following random process: at each step, a uniformly random number in $[n+k]$ is drawn, and Z returns the number of steps needed for n distinct numbers to be draw. Note that Z statistically dominates Y, because every time a man proposes to a woman, he is at least as likely to propose to an unmatched woman as the coupon collect step is to select an unselected outcome.

Now, we have $Z = \sum_{i=1}^{n} Z_i$, where Z_i is the number of steps needed before the ith distinct number is draw. Observe that Z_i is a geometric random variable with success probability $\frac{n+k-i}{n+k}$. Thus:

$$\begin{aligned}
\mathbb{E}\left[Y\right] &\leq \mathbb{E}\left[Z\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[Z_i\right] \\
&= \sum_{i=1}^{n} \frac{n+k}{n+k-i} \\
&= (n+k) \sum_{i=k}^{n+k} \frac{1}{i} \\
&= (n+k)[H_{n+k} - H_k] \\
&= \Theta\left((n+k)[\log(n+k) - \log(k)]\right)
\end{aligned}$$

Which gives exactly the desired bound on $\mathbb{E}\left[R_{men}(MOSM(P))\right] = \mathbb{E}\left[Y/n\right]$. ☐

We can use the above claim to get an informal handle on the expectation of $R_{women}(MOSM(P))$ as well. In expectation, $O\left((n+k)\log\left(\frac{n+k}{k}\right)\right)$ proposals are made, so any fixed woman w receives an average of $O\left(\log\left(\frac{n+k}{k}\right)\right)$ proposals. The rank of the men proposing to w is essentially uniformly distributed on $[n]$. The minimum of h uniformly distributed random variables with expectation x has expectation x/h. Thus, the best-ranked man proposing to w has rank $\Omega\left(n/\log\left(\frac{n+k}{k}\right)\right)$ on average.

The previous paragraph can be made formal. A sequence of classical works (in order, [Wil72], [Knu97], [Pit89], [Pit92]) got quite a good handle on the asymptotic behavior of balanced matching markets. Performing a full analysis can get quite complicated, especially when counting the *number* of distinct stable matchings[4].

6 Unbalanced Random Markets: One Exposition

With a little effort, you can extend coupon-collector type arguments to get results in the unbalanced case. This next claim is adapted from [IM05]:

Claim 6.1. *Consider some matching market P (i.e. a set of preferences for men and women) and fix a woman w^* who is matched under any some stable matching for P. Construct preference sets $P_n, P_{n-1}, \ldots, P_1$ where in P_i we have w^* "truncate her list" after place i, that is, w^* reports her top i choices and deems all other matches unacceptable. Let m_i denote the match of w^* under $MPDA(P_i)$ (or write $m_i = \emptyset$ if w^* is unmatched) and let $L = (m_n, m_{n-1}, \ldots, m_1)$.*

We have the following:

1. *For each i, μ_i is a stable matching for P_i with $\mu_i(w^*) \neq \emptyset$ if and only if μ_i is a stable matching for P.*

2. *L is a list containing all the stable matches of w, possibly with repetition and possibly terminating in a string of $\emptyset$s (which do not represent stable matches).*

3. *L is ranked from worst to best according to $\prec_w$.*

Proof. (1, $\Rightarrow$) If $\mu_i(w^*) = \emptyset$, then we immediately know that μ_i is not stable for P, by the rural hospital theorem (claim 2.4 suffices). Also, if μ_i is not stable for P_i, then any blocking pair in P_i will also be a blocking pair in μ_i, so μ_i will not be stable for P.

(1, $\Leftarrow$) Suppose w^* is matched in μ_i, and suppose for contradiction that (m, w) is a blocking pair for μ_i under P. All agents have identical preferences in P and P_i except for w^*, so we must have $w = w^*$. But because w^* is matched under P_i, we know $m \succ_{w^*} \mu_i(w^*)$ if and only if $m \succ_{w^*}^{(i)} \mu_i(w^*)$ (where $\succ_{w^*}^{(i)}$ denotes w^*'s preferences in P_i). Thus, (m, w) is also blocking for μ_i under P_i, a contradiction.

(2, 3) As in the case of full-length preference lists, $MPDA(P_i)$ will return a stable matching for P_i, and it will assign w^* to the worst stable partner w^* has under preferences

[4] This is likely one reason that [AKL17] uses different statistics to measure the size of the core, namely, the fraction of agents with multiple distinct partners. For many applications, such as the strategic implication of the size of the core, this is the more important statistic anyway.

P_i, provided one exists. By part 1, the stable partners of w^* under P_i are exactly the stable partners of w^* under P which are ranked better than place i.

Now, induct on j, with the inductive hypothesis being that $(m_n, m_{n-1}, \ldots, m_j)$ contains all stable partners ranked above j by woman w^* (in order according to $\prec_{w^*}$). The base case is simply the fact that m_n is the worst stable partner of w^* (claim 2.3). Now, for the inductive step, we have two cases. If m_j is still on the preference list of w^* in P_{j+1}, then μ_j will still be a stable match under preferences P_{j+1}. Again by claim 2.3, woman w^* will still be matched to m_j under preferences P_{j+1}. On the other hand, if m_j is no longer acceptable to w^* in P_{j+1}, then m_{j+1} will be the stable partner of w^* which is worst-ranked according to $\prec_{w^*}$ (or $\emptyset$ if w^* has already reached her best stable match) by the same logic. This finishes the proof of the inductive step and of claims 2 and 3.

□

We get this immediate corollary:

Corollary 6.2. *Suppose that when a woman w^* truncates her preference list after place i, she is unmatched by MPDA. Then w^* has no stable parters ranked better than i.*

We are now ready to prove the main result of this section. We focus on the case of n men and $n+1$ women, for simplicity. I'm still working on making this bit fully formal.

Claim 6.3. *Consider a random matching market with n men and $n+1$ women. Fix some woman w^*. With probability approaching 1 as $n \to \infty$, w^* has no stable partners ranked better than $\Omega(n/\log n)$.*

(Proof Sketch). Suppose w^* truncates her preference list at place $cn/\log n$ for some constant c. Denote this preference list by P^*. By corollary 6.2, w^* has a stable partner ranked better than $cn/\log n$ if and only if w^* is matched under $MPDA(P^*)$. This occurs if and only if w^* receives a proposal from a man ranked better than $cn/\log n$. But when a man proposes to w^*, his rank by w^* is uniformly distributed (over those ranks w^* has not yet seen in the algorithm). To proceed, we upper bound the number of proposals that w^* likely receives, and thus the probability that she gets a proposal ranked higher than $cn/\log n$.

Let Y denote the random variable giving the number of proposals w^* sees during the run of $MPDA(P^*)$. First, note that $MPDA(P^*)$ will certainly terminate if all women other than w^* have been proposed to. Thus, Y is statistically dominated by the following variation on a coupon collector random variable: define a process which at each step selects a number uniformly at random from $[n+1]$. The process terminates when each number in $[n]$ has been selected, and Z outputs the number of times $n+1$ was selected. We have Y statistically dominated by Z.

Now, because the standard coupon collector random variable has expectation $O(n \log n)$, this modified coupon collector should have expectation $O(\log n)$. Thus, in expectation, the best ranked man who proposes to w^* has rank $\Omega(n/\log n)$.

□

7 [AKL17] and MOSM to WOSM Conversion

For this section, we always assume $|\mathcal{M}| = |\mathcal{W}| + k$ for $k > 0$. The main algorithm used in the analysis of [AKL17] is given here in Algorithm 2. To understand this algorithm fully, I've actually implemented it at `https://github.com/ClathomasPrime/CompetitiveStableMatching`. It is presented here in a more "imperative programming" style than that given in [AKL17]

Finally, [AKL17] shows that the core of an unbalanced market is probably small by showing that algorithm 2 probably terminates quickly. Thus, there cannot be a big difference between the man-optimal and the woman optimal stable matches: most agents have unique stable partners, and the average ranks for agent's partners is about the same in all matchings. This is a long, involved probabilistic argument, but is terribly difficult to carry out. The result is:

Theorem 7.1 ([AKL17], Appendix B)**.** *Consider any sequence of random matching markets with n men and $n+k$ women (for any $k = k(n) \geq 1$). For any $\epsilon > 0$, with probability that approaches 1 as $n \to \infty$, each of the following events hold:*

1. *Define $s_k(n) := \left(\frac{n+k}{n}\right) \log\left(\frac{n+k}{k}\right)$. Then for every stable matching μ,*

$$R_{men}(\mu) \leq (1+\epsilon) s_k(n) \qquad\qquad R_{women}(\mu) \geq \frac{n}{1 + (1+\epsilon) s_k(n)}$$

2. *Less than a $1/\sqrt{\log n}$ fraction of the men and women have multiple stable partners. In particular, this fraction approaches 0 as $n \to \infty$. Moreover, the total number of stable partners is at most $n + n/\sqrt{\log n}$.*

3. *There is little difference in ranking between the MOSM and the WOSM:*

$$\frac{R_{men}(WOSM)}{R_{men}(MOSM))} \leq 1 + \epsilon \qquad\qquad \frac{R_{women}(WOSM)}{R_{women}(MOSM))} \geq 1 - \epsilon$$

Comparing this to the results from section 3, we see that when imbalance is present, men *almost always* do as well in *any* stable matching as they do in the man-optimal stable matching. Two special cases are worth pointing out:

Corollary 7.2 (2.2 in [AKL17])**.** *Suppse $k(n) = 1$ for each n. Then for every $\epsilon > 0$, with probability approaching 1 as $n \to \infty$, we have*

$$R_{men}(\mu) \leq (1+\epsilon) \log n \qquad\qquad R_{women}(\mu) \geq (1-\epsilon) \frac{n}{\log n}$$

for every stable matching μ.

Corollary 7.3 (2.3 in [AKL17])**.** *Suppse there exists a $\lambda > 0$ with $k(n) = \lambda n$ for each n. Let $\kappa = (1+\lambda)\log(1 + 1/\lambda)$. Then for every $\epsilon > 0$, with probability approaching 1 as $n \to \infty$, we have*

$$R_{men}(\mu) \leq (1+\epsilon)\kappa \qquad\qquad R_{women}(\mu) \geq (1-\epsilon) \frac{n}{1+\kappa}$$

for every stable matching μ.

In a follow-up paper [Pit19], many of these results are shown to be sharp, and the analysis is extended to count the average number of stable matchings in an unbalanced market. Interestingly, [Pit19] is not able to improve the rate at which the fraction of agents with multiple stable partners tends to zero as $n \to \infty$, even though the arguments in section 4 seem to suggest that this fraction might be about $1/\log n$.

It's worth noting that algorithm 2 is almost identical to the algorithm *minimal-differences* described in [GI89] (a book which fully fleshes out ideas initiated in [IL86]). The purpose of algorithm *minimal-differences* is to traverse the matching lattice in a maximal chain from the MOSM to the WOSM, and along the way identify "the rotation poset" (i.e. the poset for which the matching lattice is the collection of downwards-closed sets of the poset, which is guaranteed to exist by Birkhoff's representation theorem). The only difference is that algorithm 2 does not explicitly identify the "rotations" (i.e. simple improvement cycles). [AKL17]'s careful analysis of this algorithm for unbalanced random markets essentially says that the rotation poset, and consequentially the stable matching lattice, is typically small (in a few very precise ways).

References

[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.

[DF81] E. L. Dubins and A. D. Freedman. Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, 88:485–494, 08 1981.

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

[GS85] David Gale and Marilda Sotomayor. Ms. machiavelli and the stable matching problem. *The American Mathematical Monthly*, 92(4):261–268, 1985.

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

[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.

[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.

[Knu97] Donald Knuth. *Stable Marriage and its Relation to Other Combinatorial Problems (Translation by Martin Goldstein of a 1977 French publication)*, volume 10 of *CRM Proceedings and Lecture Notes*. 1997.

[Pit89] Boris Pittel. The average number of stable matchings. *SIAM J. Discret. Math.*, 2(4):530–549, November 1989.

[Pit92] Boris Pittel. On likely solutions of a stable marriage problem. *Ann. Appl. Probab.*, pages 358–401, 1992.

[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.

[Wil72] L. B. Wilson. An analysis of the stable marriage assignment algorithm. *BIT Numerical Mathematics*, 12(4):569–575, Dec 1972.