Equitable stable matchings under modular assessment*

AHMET ALKAN† KEMAL YILDIZ‡

August 24, 2023

Abstract

We propose a framework for addressing issues of equity and social welfare in the stable matching model. We first establish an equivalence between an ordinal condition and modular optimization on the lattice of stable matchings. This equivalence charts out a domain where equity or welfare criteria additively separate over agents' *attainable* mates and appear as weights in optimization. We call the ordinal condition *convexity* and the domain *modular*. *Convexity* requires stable "mixtures" of matchings in a solution to also be in the solution. We next propose a novel class of equitability criteria called *equity undominance* and characterize the modular stable matching rules that are equity undominated. It follows from our results that the modular stable matching rules provide for clear testable implications and a wide range of specifications allowing efficient optimization.

Keywords: Equity, attainability, lattice, rotations, identification, modular optimization.

*We thank Battal Doğan, Serhat Doğan, Faruk Gul, Tarık Kara, Deniz Savas, seminar participants at Bristol University, Queen Mary University, City University of London, University of Alicante, University of Maryland, participants of several conferences, and anonymous referees for valuable comments and suggestions. Kemal Yıldız's research is supported by the BAGEP Award of the Science Academy in Turkey and the Marie Curie International Reintegration Grant (# 837702) within the European Community Framework Programme. An extended abstract published in the *24th ACM Conference on Economics and Computation Proceedings*. First posted version: February 8, 2022.

†Department of Economics, Sabancı University; alkan@sabanciuniv.edu.tr

‡Department of Economics, Bilkent University and Princeton University; kemal.yildiz@bilkent.edu.tr

Contents

1 Introduction

An important feature of matching markets is that there typically exist many *stable* matchings. These matchings have a remarkable orderliness property in two-sided markets. They form a lattice according to the group preferences of one side that is opposite to the group preferences of the other side. The two extremal matchings, optimal for one side pessimal for the other, bear extreme inequity. Nonetheless, since the seminal work of Gale & Shapley (1962), research and applications in the area mostly involved the extremal matchings and much less so the middle of the stable set where inequity may be resolved. This is partly because the optimal stable matching has proved very useful in centralized market applications on account of the strategyproofness and algorithmic properties it has. It is also because the "middle" has proved challenging definitionally as well as computationally. Besides which, studying equitable matchings is economically relevant, since decentralized markets tend to converge to the middle of the stable set (Echenique, Robinson-Cortes & Yariv 2022).[1]

We see two lines of approach for locating equitable matchings in the middle. One line features ordinal criteria only; a prominent example is the *median stable matching* (Teo & Sethuraman 1998). The other employs one or another value function over the stable set whose minima are deemed equitable; an example is the *sex-equal stable matching* (Gusfield & Irving 1989) named after the function that evaluates the difference between the total rank achievements of the two sides. These solutions substantially disagree with each other, moreover, nearly all pose computational difficulty. We hold that the disparity among these solutions, and the breadth of possibilities in general, call for a foundational framework to address issues of social welfare and equity in the stable matching model. Here we propose *modular stable matching rules* to form such a framework. Our approach employs both value functions and ordinal criteria with a one-to-one correspondence between the two. The modular stable matching rules turn out to be analytically tractable, easy to identify, and rich enough to implement a wide range of objectives.

[1]They conduct experiments in which agents freely interact with (make offers to) each other to form matchings. Their three main findings are motivating for our study. First, most agents end up with their "median attainable (stable) partner" when they have three attainable (stable) partners. Second, the median outcome is still the most common even if only one side of the market can make offers. Third, the stable matchings that are formed depend on how preferences are represented in cardinal terms.

To introduce the framework, consider a society consisting of equal numbers of *men* and *women*. A matching *problem* is a preference profile in which each agent has preferences over the opposite party. A *matching* uniquely assigns each man to a woman. Our primitive objects are *matching rules* that associate a nonempty set of matchings with each problem. The set of associated matchings can be thought of as the matchings assigned positive probability in a lottery, a shortlist from which a final choice is to be made, or the matchings used in a rotation scheme that, for instance, specifies the periodical job assignments or how to use common-pool resources.

The central robustness criterion for a matching is *stability*, which requires that there is no unmatched man-woman pair who prefer each other to their assigned mates. We require that a matching rule chooses only stable matchings. Additionally, we impose *invariance under stability*, which requires the matchings chosen from two problems to be the same unless these problems induce different stable matchings. We discuss this restriction in connection to the existing approaches to equitability. The main restriction we impose on a matching rule is *modularity* that requires the chosen matchings be the ones that optimize a *modular* objective function. We refer to the matching rules that satisfy the three conditions as *modular stable matching rules*.

To describe the intuition behind modular stable matching rules, the critical notion is *attainable mates*. A man and a woman are *attainable* for each other in a given problem if a stable matching exists in which they are matched to each other. In Proposition 1, we show that an objective function is modular if and only if the value of a stable matching is the sum of the social value of matching each agent with their mate in the matching. Thus, under the premiss that unattainable agents should not matter, a modular objective function represents an equity or welfare criterion that is additively separable over agents attainable mates.

We provide several examples and results demonstrating that a wide range of objectives can be implemented through modular stable matching rules. For analytical tractability, Proposition 1 and the results of Picard (1976) and Irving, Leather & Gusfield (1987) imply that computing a modular stable matching rule boils down to finding the *minimum cuts* in a properly defined *flow network*. The latter problem has been extensively studied and is known to be solvable efficiently.[2]

[2]See for example Picard & Queyranne (1980), Irving, Leather & Gusfield (1987), and Henzinger, Noe, Schulz & Strash (2020).

Our first main contribution consists of two characterization results that provide a fairly general way to think about the issues of social welfare and equity in the stable matching model. These characterizations reveal the ordinal content of optimizing a modular function under the constraint of *stability*. Thus, we provide testable axioms to identify if a society's observed choice of stable matchings complies with the optimization of a modular function. These axioms are tailored to facilitate identification of the underlying objectives in matching markets from the observable choices.[3] An analogous result is Afriat (1967)'s characterization of demand correspondences that are consistent with concave optimization. On the technical side, we deepen and use the connection between the stable matchings lattice and the *rotations poset* (Irving & Leather 1986), which is used in computer science to design efficient matching algorithms.[4]

Our first characterizing axiom, *convexity*, states that if a rule chooses two matchings, and there is a stable matching that assigns each agent to one of their mates in the chosen matchings, then the rule should also choose this stable matching. The second characterizing axiom imposes a restriction by linking choices in two different problems. Suppose we transform agents original preferences by moving their mates in the chosen matchings to the top of their preferences, keeping the relative rankings elsewhere. Our *independence of irrelevant rankings* axiom requires that if a matching is stable under both the original and the transformed preferences, then it should be chosen.

In the second part of the paper, we propose a new equity notion based on the notion of the "median attainable mate". A man and a woman are *attainable* for each other in a given problem if a stable matching exists in which they are matched to each other. Next, for each agent, from among their attainable mates, consider the one(s) with (a) *median rank*.[5] *The median attainable mate* is the more preferred attainable mate with a median rank. *Equity undominance* requires that if a matching is chosen, then there is no other stable matching in which each agent's mate is same or closer to their median attainable mate. Section 4 presents a simple problem in which a stable matching assigns all the agents to their unique median attainable mates. However,

[3] Echenique, Robinson-Cortes & Yariv (2022) present an instance of decentralized matching markets in which agents' cardinal preferences are observable, which we identify from the chosen matchings in a given problem.

[4] See for example Cheng, McDermid & Suzuki (2016) and our Example 3.

[5] If there is an even number of attainable mates for an agent, then there are two such mates.

several stable matching rules from the literature fail to choose this matching, thus violating *equity undominance*. In Theorem 2, we characterize modular stable matching rules that satisfy *equity undominance*. Finally, we present a modular stable matching rule that satisfies *equity-undominance,* and can be computed in polynomial time.

1.1 Related literature

Connections to matching theory: Existing foundational studies in the stable matching model have focused on extremal stable matchings, including Balinski & Sönmez (1999), Ehlers & Klaus (2006), and Kojima & Manea (2010) who offer different characterizations of the Gale-Shapley student-optimal stable matching rule. An axiomatic approach to the problem of fair algorithms was presented by Masarani & Gokturk (1989). However, their approach concentrates on the algorithm, not the resulting matchings, and concludes with an impossibility result.

Equitability in marriage markets is a long-standing matter. The fact that the fortunes of men and women across stable matchings are polar opposites in an orderly way points at meeting in the middle, but where in the middle resists specification. There is now a list of solutions that have been put forward, some on geometric ground, others optimizing a social objective. These solutions that we will critically review have more frequently been looked at on computational aspects.[6] In contrast, our approach is "principles-based" in that we first propose principles, then characterize the stable matching rules in terms of the underlying objectives that comply with these principles.

Connections to decision theory: We present characterizations for modular stable matching rules that provide for testable foundations. Modularity properties of agents' preferences were investigated by Kreps (1979) in decision-theory literature and by Milgrom & Shannon (1994) in the monotone comparative statics literature. Chambers & Echenique (2008) clarify the connection between these two approaches. The closest results to our Theorem 1 are due to Kreps (1979)

[6]Here, Klaus & Klijn (2006) can be viewed as an exception. They introduce the *procedural fairness* notion that labels a probabilistic stable matching rule as fair if each agent has the same probability to move at a certain point in the procedure. However, their approach does not provide any criterion for fairness or equitability of a stable matching.

and Chambers & Echenique (2009), who provide representations for modular preferences over lattices under *monotonicity*. However, violation of *monotonicity* is our departure point, since a stable matching rule that satisfies *monotonicity* would choose an extremal matching. As demonstrated in our proofs, here, the neat geometric structure of the stable matchings paves the way for our results in the absence of *monotonicity*. Another key modeling difference is that our primitives are not agents' preferences but matching rules that can be thought of as "choice rules" over the set of stable matchings.

2 Modular Stable Matching rules

2.1 Stable matching rules

Let M be a set of n men and W be a set of n women. Each $m \in M$ has preferences over W and each $w \in W$ has preferences over M. For $N = M \cup W$, preferences of each agent $i \in N$ is represented by a **strict ordering**, denoted by $\succ_i$, which is a *complete*, *transitive*, and *asymmetric* binary relation over the members of the opposite side. Let $\mathcal{P}_i$ denote the set of all possible preference relations for agent i, and $\mathcal{P}$ denote the set of all preference profiles $\times_{i \in N} \mathcal{P}_i$. We denote a generic preference profile by $\succ$.

A **matching** is a one-to-one function $\mu : M \cup W \to M \cup W$ such that for each $(m, w) \in M \times W$, we have $\mu(m) \in W$, $\mu(w) \in M$, and $\mu(m) = w$ if and only if $\mu(w) = m$. A matching μ is **stable** at a problem $\succ \in \mathcal{P}$ if there is no **blocking-pair** $(m, w) \in M \times W$ such that $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$. Let $\mathcal{S}(\succ)$ denote the set of all stable matchings at a given preference profile (problem) $\succ \in \mathcal{P}$. We often use $\mathcal{S}$ instead of $\mathcal{S}(\succ)$ if the problem that is referred to is clear from the context. Let $\rhd_M$ denote the **men-wise better than** relation over $\mathcal{S}$, which is defined as follows: For each distinct $\mu, \mu' \in \mathcal{S}$, $\mu \rhd_M \mu'$ if and only if for each $m \in M$, $\mu(m) \succ_m \mu'(m)$ or $\mu(m) = \mu'(m)$. The **women-wise better than** relation $\rhd_W$ is defined similarly. A **matching rule** is a mapping π that associates each problem $\succ \in \mathcal{P}$ with a nonempty set of matchings $\pi(\succ)$.

Definition. *A **stable matching rule** π is a matching rule that satisfies the following two conditions:*

Stability: *For each problem $\succ \in \mathcal{P}$, $\pi(\succ) \subset \mathcal{S}(\succ)$.*

Invariance under stability: *For each $\succ, \succ' \in \mathcal{P}$, if $\mathcal{S}(\succ) = \mathcal{S}(\succ')$ then $\pi(\succ) = \pi(\succ')$.*

Stability requires the matchings chosen for each problem to be stable. *Invariance under stability* requires the matchings chosen for two different problems be the same unless these problems induce different sets of stable matchings. Intuitively, the set of stable matchings of each problem should provide the relevant information for a stable matching rule. Examples 4-6 demonstrate that for a stable matching rule to satisfy *invariance under stability*, the critical notion is "attainable mates". A man and a woman are *attainable* for each other in a given problem if a stable matching exists in which they are matched to each other.

The existing matching rules can roughly be classified into two groups. The rules in the first group choose the stable matchings optimizing an explicitly given objective function, which are interpreted as a measure of social welfare or fairness–such as the utilitarian objective (the sum of individual utilities).[7] A common theme in these rules is that the objective function to be optimized is formulated via agents' rankings over each other, independent of whether them being *attainable* or not. Thus, all fail to satisfy *invariance under stability*. To see this, let $\succ$ be a problem and consider the problem $\succ'$ obtained from $\succ$ such that agents' *attainable mates* are moved to the top of their preferences by preserving the relative rankings within attainable and unattainable mates. It is easy to verify that the associated set of stable matchings remains the same. However, since $\succ$ and $\succ'$ can be distinct, optimizing an objective function based on agents' rankings might result in different stable matchings. As demonstrated for the sex-equal stable matchings in Example 6, to remedy the violation of *invariance under stability*, it is sufficient to replace the use of agents' rankings with their "attainable rankings".

The rules in the second group are built on the geometric structure of stable matchings. That is, for each pair of stable matchings μ and μ', consider $\mu \vee \mu'$ ($\mu \wedge \mu'$) that maps each man to

[7]For other examples, one can count *minimum regret stable matchings*, *egalitarian stable matchings*, *minimum weight stable matchings*, *sex-equal stable matchings, rank maximal stable matchings, and balanced stable matchings.* For the related definitions, we refer the reader to Manlove (2013, Chapter 1.3), Gusfield & Irving (1989), and the references therein.

his best (worst) mate among the women he is matched to at μ or μ'; it turns out that both $\mu \vee \mu'$ and $\mu \wedge \mu'$ are stable matchings as well, which in particular implies that the pair $\langle \mathcal{S}, \rhd_M \rangle$ forms a **distributive lattice**.[8] Next, we present three examples whose formulations are based on the structure of the stable matchings lattice, satisfy invariance *under stability*, thus exemplifying stable matching rules.

Example 1 (**Median stable matching**)**.** Let $\succ$ be a problem with K stable matchings. For each man m, arrange his mates from these K stable matchings from his most preferred mate to his least preferred one. Let $w^k(m)$ denote the k*-th* woman in this sorted list, where a woman is counted as many times as she occurs as a match in different stable matchings. For each $k \in \{1, \ldots, K\}$, define $\mu^k : M \cup W \rightarrow M \cup W$ such that $\mu^k(m) = w^k(m)$ for each $m \in M$. Teo & Sethuraman (1998) show that μ^k is a stable matching. Then, they define the *median stable matching(s)* as $\mu^{(K+1)/2}$ when K is odd and $\mu^{K/2}$ and $\mu^{(K/2)+1}$ when K is even.[9]

Example 2 (**Median of the stable matchings lattice**)**.** Consider the undirected graph associated with the stable matchings lattice $\langle \mathcal{S}, \rhd_M \rangle$, in which each $\mu \in \mathcal{S}$ is a vertex and for each $\mu, \mu' \in \mathcal{S}$, μ and μ' are *adjacent* if there is no $\mu'' \in \mathcal{S}$ with $\mu \rhd_M \mu'' \rhd_M \mu'$.[10] For each $\mu, \mu' \in \mathcal{S}$, the *distance* between μ and μ' is the length of (the number of edges on) the shortest path (*geodesic*) between μ and μ' in this graph. Cheng (2010) analyzes stable matchings which are the *medians of the stable matchings lattice* whose total distance from all other stable matchings is the least. She shows that if n is odd, then the median stable matching is the unique median of the stable matchings lattice. If n is even, then a stable matching μ is a median of the stable matchings lattice if and only if μ is between the median stable matchings according to $\rhd_M$, i.e. $\mu^{(K/2)+1} \rhd_M \mu \rhd_M \mu^{K/2}$.

Example 3 (**Center stable matching**)**.** Cheng, McDermid & Suzuki (2016) formulate *center stable matching(s)* as the one(s) whose maximum distance (as described in Example 2) from any other stable matchings is the least. They characterize all center-stable matchings and show that a specific one center-stable matching can be found in polynomial time.

[8] Knuth (1976), pp. 92-93, attributes the discovery of this lattice structure to J. H. Conway.

[9] The existence of (generalized) median stable matchings has been studied in other settings, including: one-to-one matching with wages (Schwarz & Yenmez 2011), the college admissions model with responsive preferences (Sethuraman, Teo & Qian 2006), the roommates problem (Klaus & Klijn 2010), many-to-many matching markets with contracts (Chen, Egesdal, Pycia & Yenmez 2016).

[10] This graph is called the (undirected) Hasse diagram of $\langle \mathcal{S}, \rhd_M \rangle$.

2.2 Modularity

The main structural restriction we impose on a stable matching rule is *modularity*. Modularity requires optimizing an explicit assessment function restricted in a specific way via the lattice structure of stable matchings. That is, for a problem $\succ$, let $F : \mathcal{S}(\succ) \to \mathbb{R}$ be an **assessment function** attaching a value $F(\mu)$ to each stable matching $\mu \in \mathcal{S}(\succ)$. Then, F is **modular** if for each $\mu, \mu' \in \mathcal{S}(\succ)$,

$$F(\mu) + F(\mu') = F(\mu \vee \mu') + F(\mu \wedge \mu'). \tag{1}$$

Note that (1) can be rewritten as $F(\mu) - F(\mu \wedge \mu') = F(\mu \vee \mu') - F(\mu')$. Then, the simple intuition behind modularity is as follows. The change from $\mu \wedge \mu'$ to μ corresponds to a group of men being matched to better women; the effect of this change should be the same if the change was made while another group of men was matched to better women (at μ') compared to their mates at $\mu \wedge \mu'$.

Definition. Let π be a stable matching rule. Then, π is **modular** if for each problem $\succ$, there exists a modular $F : \mathcal{S}(\succ) \to \mathbb{R}$ such that $\pi(\succ)$ is the set of stable matchings that minimize F, i.e. $\pi(\succ) = \text{argmin}_{\mu \in \mathcal{S}(\succ)} F(\mu)$.

As for a closely related class of assessment functions, consider the ones represented in an additively separable form $\sum_{i \in N} F_i(\mu(i))$. This representation renders the direct interpretation that the total value of a stable matching is obtained by adding $F_i(\mu(i))$ for each agent i, which assesses the social value of matching agent i with $\mu(i)$. For example, imagine that a matching determines the partnership between a senior and junior employee or two teams running a joint project at the intersection of their areas of expertise. Then, the social value of matching agent i with $\mu(i)$ might be determined by the productivity of agent i when matched with $\mu(i)$ represented by $F_i(\mu(i))$. This formulation disallows complementarities, since the social value of matching agent i with $\mu(i)$ is the same regardless of other agents' matches.

It follows from the findings of Picard (1976) and Irving, Leather & Gusfield (1987) that a stable matching optimizing a given additively separable assessment function can be found efficiently. In Proposition 1, we show that an assessment function is modular if and only if it can be represented as the sum of *individual assessment functions* defined for each agent over the set of their attainable mates. Formally, let $\succ$ be a given problem, then a man m and a woman w are **attainable** for each other if $\mu(m) = w$ for some $\mu \in \mathcal{S}(\succ)$. For each $i \in N$, let $A_i(\succ)$ denote the set of attainable agents for agent i. As usual, we will use A_i instead of $A_i(\succ)$ if the problem that is referred to is clear from the context.

Proposition 1. *Let $\succ \in \mathcal{P}$ be a problem and $F : \mathcal{S}(\succ) \to \mathbb{R}$ be an assessment function. Then, F is modular if and only if for each $i \in N$, there exists $F_i : A_i(\succ) \to \mathbb{R}$ such that $F(\mu) = \sum_{i \in N} F_i(\mu(i))$ for each $\mu \in \mathcal{S}(\succ)$.*

Here, we prove the if part of the statement. Suppose that for each $\mu \in \mathcal{S}(\succ)$, we have $F(\mu) = \sum_{i \in N} F_i(\mu(i))$, where $F_i : A_i \to \mathbb{R}$. To see that π is modular, note that for each $i \in N$ and $\mu, \mu' \in \mathcal{S}(\succ)$, $\{(\mu \vee \mu')(i), (\mu \wedge \mu')(i)\} = \{\mu(i), \mu'(i)\}$. Therefore, $\{F_i((\mu \vee \mu')(i)), F_i((\mu \wedge \mu')(i))\} = \{F_i(\mu(i)), F_i(\mu'(i))\}$. It follows tha t F_i is modular for each agent i, thus F is modular since it is the sum of these modular functions. The proof of the only if part uses the *rotations poset* presented as Lemma 4 in Section 6.1.

2.3 Examples

We present several examples to demonstrate the relevance, generality, and possible limitations of the modular stable matching rules. These rules are based on aggregating agents' **attainable rankings**. That is, for each m and w who are attainable for each other, $Rank_m^A(w)$ ($Rank_w^A(m)$) is the rank of w (m) in $\succ_m |_{A_m}$ ($\succ_w |_{A_w}$), which is obtained by restricting $\succ_m$ ($\succ_w$) to the women (men) who are attainable for m (w).

Example 4 (Maximizing total attainable ranks). In the vein of utilitarian welfare measures, it may be reasonable to evaluate each stable matching according to the sum of agents' attainable ranks in the matching. That is, for each problem $\succ$, let $\pi(\succ)$ be the set of stable matchings maximizing $\sum_{mw \in \mu}(Rank_m^A(w) + Rank_w^A(m))$. In Lemma 1 of Section 6.1, we show that this sum is constant among all stable matchings, and therefore does not differentiate any stable matching from the others.

Example 5 (Minimizing total spread from the ideals). For each agent $i \in N$, from among the agents who are attainable for i, let $I(i)$ be the *ideal* partner for i in the sense that assigning i to $I(i)$ makes agent i reach the welfare level that is found ideal for them. We allow two different agents to have the same ideal partner. For a given stable matching μ, we can measure the **spread from the ideal** for agent i by $|Rank_i^A(\mu(i)) - Rank_i^A(I(i))|$. Then, consider the stable matching rule π choosing the set of stable matchings minimizing the *total spread from the ideals*. That is, for each problem $\succ$, let $\pi(\succ)$ be the set of matchings minimizing $\sum_{i\in N} |Rank_i^A(\mu(i)) - Rank_i^A(I(i))|$. It directly follows from Proposition 1 that π is modular.

Example 6 (Minimizing the difference between total attainable ranks). As a counterpart of the *sex-equal stable matchings* (Gusfield & Irving 1989), consider the stable matching rule π that chooses the set of *attainable sex-equal stable matchings* minimizing the absolute value of the difference between each side's total attainable ranks. That is, for each problem $\succ$, let $\pi(\succ)$ be the set of matchings minimizing $|\sum_{m\in M} Rank_m^A(\mu(m)) - \sum_{w\in W} Rank_w^A(\mu(w))|$. We will see that this stable matching rule is not modular.

3 Axioms and characterizations

3.1 Convexity

We present two characterizations for modular stable matching rules revealing the ordinal content and testable implications of optimizing a modular function over stable matchings lattice. Our first axiom, *convexity*, requires that for a given pair of matchings that are chosen by the rule, if one can form a "mixture" stable matching by assigning agents to one of their mates in the given matchings, then this newly formed matching should be chosen as well.

Convexity: For each problem $\succ \in \mathcal{P}$, if $\mu', \mu'' \in \pi(\succ)$ and there exists $\mu \in \mathcal{S}(\succ)$ such that $\mu(m) \in \{\mu'(m), \mu''(m)\}$ for each $m \in M$, then $\mu \in \pi(\succ)$.

Two familiar examples of *mixtures* are the join and meet of a pair of matchings. It follows that if π satisfies *convexity*, then $\pi(\succ)$ is a sublattice of $\mathcal{S}(\succ)$. On the other hand, *convexity* is weaker than requiring that all the stable matchings between the chosen matchings–according to the men-wise better than relation–are chosen, which is also referred to as convexity in the lattice theory literature. To see this, consider a stable matching rule π such that for each problem $\succ$, only the extremal matchings are chosen whenever there is no other $\mu \in \mathcal{S}(\succ)$ such that $\mu(m) \in \{\mu^M(m), \mu^W(m)\}$ for each $m \in M$ (see the problem in Example 7); and chooses all of the stable matchings, otherwise. Although π satisfies our *convexity*, it clearly violates the latter requirement. It is worth emphasizing that *convexity* requires a matching that is a mixture of two chosen stable matchings be chosen only if this mixture matching is also stable.[11] Next, we present our first characterization result that establishes the equivalence between *modularity* and *convexity*.

Theorem 1. *Let π be a stable matching rule. Then, π is modular if and only if π satisfies convexity.*

Proof. Please see Section 7.1. □

Before we sketch the proof for the if part, we revisit the stable matching rules from the literature discussed in Section 2.1. Using our Theorem 1, we can easily check whether these rules are modular. It turns out that the matching rule that chooses the median(s) of the stable matching lattice (Example 2) is modular since it is *convex*. To see this, recall that this rule chooses the unique median stable matching or all the stable matchings that are between the median stable matchings according to the men-wise better than relation. On the other hand, the following example demonstrates that the rules presented in Examples 1,3 and 6 fail to satisfy *convexity*.

Example 7. Consider the problem with eight agents whose preferences are represented by the table in Figure 1 such that each entry ij is associated with man m_i and woman w_j, the $\succ_{m_i}$-rank of w_j is written in the bottom corner, and the rank of m_i in $\succ_{w_j}$ is written in the top corner. Figure 1 also presents the associated stable matchings lattice such that each stable matching is represented as an array $[r_1, \ldots, r_4]$, where each r_i is the $\succ_{m_i}$-rank of the woman who is matched with m_i. Then, $[2222]$ and $[3333]$ are the median stable and attainable sex-equal matchings. The centers of the stable matchings lattice are $[3322]$ and $[2233]$. Thus, all three rules fail to satisfy *convexity*.

[11] The example in Section 7.4 demonstrates that every mixture of stable matchings is not necessarily stable.

M \ W	w_1	w_2	w_3	w_4
m_1	1 \ 4	4 \ 1	2 \ 3	3 \ 2
m_2	4 \ 1	1 \ 4	3 \ 2	2 \ 3
m_3	3 \ 2	2 \ 3	1 \ 4	4 \ 1
m_4	2 \ 3	3 \ 2	4 \ 1	1 \ 4

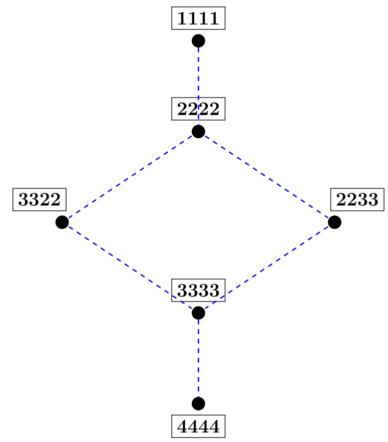


Figure 1: The problem and the associated stable matchings lattice.

3.1.1 Proof sketch: If part of Theorem 1

Let π be a stable matching rule that satisfies *convexity*. To simplify the construction of the desired modular assessment function F, we use the connection between the stable matchings and the so-called *closed* sets of *rotations*. Initially introduced by Irving (1985), rotations can be intuitively thought of as the incremental changes that transform a stable matching μ into another stable matching μ' such that there lies no other stable matching between the two (according to the men-wise better than relation). A rotation ρ *exposed in* a stable matching μ is a cyclic sequence of distinct man-woman pairs $[(m_1, w_1), (m_2, w_2) \dots, (m_k, w_k)]$ such that each m_i is matched to w_i in μ. To *eliminate* ρ and obtain μ', each man m_i in ρ is matched to w_{i+1}. A rotation ρ *precedes* another rotation ρ', if ρ must be eliminated first in order to obtain a stable matching in which ρ' is exposed. A set of rotations R is *closed* if whenever a rotation ρ is contained in R, then all the rotations that precede ρ are also contained in R. In their main result, Irving & Leather (1986) show that each stable matching μ is associated with a unique closed set of rotations R_μ. It follows from this result and simple observations made in Section 6.1, that constructing the desired F is equivalent to assigning a weight $g(\rho)$ to each rotation ρ such that $\pi(\succ)$ is the set of stable matchings that minimize $\sum_{\rho \in R_\mu} g(\rho)$.

As the second step of the proof, we introduce *hyper-rotations*, which are intuitively sets of rotations connecting the chosen stable matchings, and generically denoted by λ (see Section 6.2). Since, by *convexity*, $\pi(\succ)$ is a sublattice, let $\bar{\mu}(\underline{\mu})$ be its $\rhd_M$-best(worst) matching. Now,

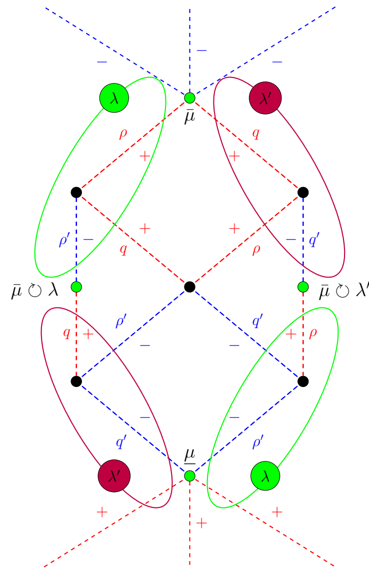


Figure 2: A demonstration of our weight assignment to the rotations, where members of $\pi(\succ)$ are lightly colored.

there can be other chosen stable matchings between $\bar{\mu}$ and $\underline{\mu}$ that are obtained by eliminating hyper-rotations. Therefore, the weights of the rotations should be assigned such that the total weight of each hyper-rotation is zero, while the total weight of each (relatively) closed set of rotations $R \subsetneq \lambda$–which corresponds to an unchosen stable matching between two chosen matching–is positive. For a fixed hyper-rotation λ, achieving this requires "preloading" in the sense that we assign positive weights to the rotations with no predecessors in λ and assign negative weights (say -1) to the rotations with no successors in λ, while assigning zero weights to all other rotations in λ.

In Figure 2, we demonstrate this construction for a straightforward case. For the general case, the inner structure of hyper-rotations, and therefore *convexity*, plays a crucial role. In Lemma 7, we show that *convexity* implies that each hyper-rotation is a *connected* set of rotations. This observation paves the way to preload the positive weights in a hyper-rotation meticulously, in that we can distribute each negative weight assigned to a rotation ρ with no successors, equally among the rotations with no predecessors and precede ρ. We complete the proof by showing that we obtained the desired weight assignments through this construction which is demonstrated in Figure 7 in Section 7.1.

3.2 Independence of irrelevant rankings

To introduce our second axiom, we need the notion of a π-*transformed problem*. For each problem $\succ \in \mathcal{P}$ and each agent $i \in N$, let $\pi_i(\succ)$ be the set of agents that i is assigned to in any matching $\mu \in \pi(\succ)$, i.e. $\pi_i(\succ) = \{\mu(i) \in N \mid \mu \in \pi(\succ)\}$. Then, the π-**transformed problem** $\succ^\pi$ is the problem obtained from $\succ$ such that for each agent $i \in N$, each member of $\pi_i(\succ)$ is moved to the top of agent i's preferences by preserving the relative rankings elsewhere.

For each $\mu \in \mathcal{S}(\succ)$, if $\mu \in \pi(\succ)$, then in transforming $\succ$ into $\succ^\pi$, for each $i \in N$, the set of agents that i prefers to $\mu(i)$ remains the same or shrinks, i.e. $\{j \mid j \succ_i^\pi \mu(i)\} \subset \{j \mid j \succ_i \mu(i)\}$. Therefore, for each $\mu \in \mathcal{S}(\succ)$, if $\mu \in \pi(\succ)$ then $\mu \in \mathcal{S}(\succ^\pi)$. Our next axiom requires the converse. That is, if a stable matching remains stable after the transformation, then it must be one of the matchings chosen by the rule in the initial problem. It follows from Proposition 2 that *convexity* and *independence of irrelevant rankings* separately characterize modular stable matching rules.

Independence of Irrelevant Rankings (IIR): For each problem $\succ \in \mathcal{P}$ and $\mu \in \mathcal{S}(\succ)$, if $\mu \in \mathcal{S}(\succ^\pi)$, then $\mu \in \pi(\succ)$.

Proposition 2. *Let π be a stable matching rule. Then, π satisfies convexity if and only if π satisfies independence of irrelevant rankings.*

Proof. Please see Section 7.2. □

To see one side of the connection between *convexity* and *IIR*, we show that if a stable matching rule π satisfies *IIR*, then π satisfies *convexity*. For a given problem $\succ$, let $\mu', \mu'' \in \pi(\succ)$ and $\mu \in \mathcal{S}(\succ)$ such that for each $m \in M$, $\mu(m) \in \{\mu'(m), \mu''(m)\}$. Since π satisfies *IIR*, we have $\pi(\succ) = \mathcal{S}(\succ) \cap \mathcal{S}(\succ^\pi)$. Therefore, to conclude that $\mu \in \pi(\succ)$ it is sufficient to show that $\mu \in \mathcal{S}(\succ^\pi)$. By contradiction, suppose there is a blocking pair (m, w). Therefore, $w \succ_m^\pi \mu(m)$ and $m \succ_w^\pi \mu(w)$. Now, since we have $\mu(m) \in \{\mu'(m), \mu''(m)\}$ and $\mu(w) \in \{\mu'(w), \mu''(w)\}$, where $\mu', \mu'' \in \pi(\succ)$, by the definition of the π-transformation, no woman is moved over $\mu(m)$ in m's preferences and no man is moved over $\mu(w)$ in w's preferences while moving from $\succ$ into $\succ^\pi$. Then, it follows from $w \succ_m^\pi \mu(m)$ and $m \succ_w^\pi \mu(w)$ that $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$, contradicting that $\mu \in \mathcal{S}(\succ)$.

4 A new class of equity notions

Studying modular stable matching rules opens new avenues through which the issues of equity can be fruitfully analyzed. To demonstrate this, let us consider the problem and its stable matchings lattice presented in Figure 3. We follow the notation used in Example 7, adding that if a pair is unattainable, then the associated entry is shadowed.

M \ W	w_1	w_2	w_3	w_4	w_5	w_6
m_1	1 / 6	2 / 5	3 / 4	4 / 3	5 / 1	6 / 2
m_2	2 / 5	1 / 6	6 / 2	3 / 4	4 / 3	5 / 1
m_3	5 / 1	6 / 2	1 / 6	2 / 5	3 / 4	4 / 3
m_4	4 / 3	5 / 1	2 / 5	1 / 6	6 / 2	3 / 4
m_5	3 / 4	4 / 3	5 / 1	6 / 2	1 / 6	2 / 5
m_6	6 / 2	3 / 4	4 / 3	5 / 1	2 / 5	1 / 6

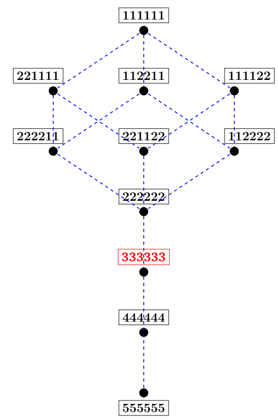


Figure 3: The problem and the associated stable matchings lattice.

Consider the stable matching in which each man is matched to his second-ranked woman and each woman is matched to her fifth-ranked man. It is easy to see that this is the unique median stable matching, the unique median of the stable matching lattice, and the unique center stable matching. However, the matching in which each man and woman is matched to their third-ranked attainable mate is reasonably more equitable, since each agent is matched to their "median attainable mate".

To formalize this intuition as a principle, for each agent i, from among the agents who are attainable for i, consider the one(s) with (a) *median rank*. Note that if there is an odd number of attainable agents for i, then there is a unique attainable agent with this property; otherwise, there are two such agents. The **median attainable mate** for agent i, denoted by med_i^A, is the (more preferred) attainable agent with the lowest attainable median rank, i.e. $Rank_i^A(med_i^A) = \lfloor |A_i|/2 \rfloor$.

Now, let i be an agent and $j, j' \in A_i$ be a pair of attainable mates for i. Agent j is **closer to** med_i^A than agent j' if $|Rank_i^A(j) - Rank_i^A(med_i^A)| < |Rank_i^A(j') - Rank_i^A(med_i^A)|$. Our following axiom requires that if a matching rule chooses a stable matching μ, then there is no other stable matching in which each agent is assigned to the same mate or someone closer to their median attainable mate compared to their mate at μ.

Equity undominance: For each problem $\succ \in \mathcal{P}$, if $\mu \in \pi(\succ)$, then there is no $\mu' \in \mathcal{S}(\succ)$ such that for each $i \in N$ with $\mu(i) \neq \mu'(i)$ we have $\mu'(i)$ is closer to med_i^A than $\mu(i)$.

In our previous example, this principle uniquely pins down the matching in which each agent is matched to their third-ranked mate. Next, we characterize the modular stable matching rules that satisfy *equity undominance*. For each agent i, the individual assessment function $F_i : A_i \to \mathbb{R}$ is **unimodal** with mode med_i^A if F_i is monotonically increasing for $med_i^A \succ_i j$ and monotonically decreasing for $j \succ_i med_i^A$.[12]

Theorem 2. *A modular stable matching rule π satisfies equity undominance if and only if for each problem $\succ \in \mathcal{P}$, $\pi(\succ)$ is the set of stable matchings that minimize $-\sum_{i \in N} F_i(\mu(i))$, where $F_i : A_i \to \mathbb{R}$ is unimodal with mode med_i^A for each $i \in N$.*

Proof. Please see Section 7.3. □

In addition to those mentioned above, the stable matching rule presented in Example 6, which chooses the set of attainable sex-equal stable matchings fails to satisfy *equity undominance*.[13]

Remark 1. The notion of *equity undominance* and Theorem 2 can be generalized by replacing median attainable mates with ideal mates as defined in Example 5. Thus, we obtain a rich class of equity notions in which ideal mates are freely specified. On the other hand, requiring that each agent is attached to a unique ideal mate can be a rather demanding feature that rules out reasonable modular stable matching rules such as the rule presented in Example 2.

[12]Put differently, F_i attains its maximum at med_i^A, and for each $j, j' \in A_i \setminus \{med_i^A\}$, we have $F_i(j) > F_i(j')$ if j' is further away from med_i^A compared to j according to $\succ_i$, i.e. $j' \succ_i j \succ_i med_i^A$ or $med_i^A \succ_i j \succ_i j'$.

[13]See the problem presented in Section 7.4, where the unique attainable sex-equal stable matching is equity dominated.

4.1 The equal weight median rule

One can consider choosing all the equity-undominated stable matchings for each problem. However, the set of equity-undominated stable matchings is turns out to be rather unstructured. This set may not be a sublattice of the stable matchings, thus failing to satisfy *convexity*.[14] This observation motivates us to present a modular stable matching rule that satisfies *equity undominance*, and can be computed in polynomial time. We obtain this rule by adding structure into the modular stable rule presented in Example 5. We view med_i^A as the ideal partner for i. Then, the **equal weight median rule** chooses stable matchings that minimize the total distance from the median. That is, for each problem, the chosen stable matchings are the ones that minimize $\sum_{i \in N} |Rank_i^A(\mu(i)) - Rank_i^A(med_i^A)|$.

This objective function can be expressed as a member of the additive unimodal family presented in Theorem 2, where for each agent i and stable matching μ, $F_i(\mu(i)) = -|Rank_i^A(\mu(i)) - Rank_i^A(med_i^A)|$. Therefore, the equal weight median rule satisfies *modularity* and *equity undominance*. As for computational efficiency, it follows from Gusfield (1987)[15] that identifying attainable mates is a polynomial task. Then, by using the findings of Irving, Leather & Gusfield (1987), one can show that a stable matching minimizing the total distance from medians can be found in polynomial time.[16] These observations extend to rules that weigh agents differently and choose stable matchings minimizing the total weighted distance from their medians. A recent study by Can, Pourpouneh & Storcken (2023)–who analyze distance functions on matchings–provide further justification for this class. The conditions they introduce in quantifying the similarity between two matchings characterize distances that scale the sum of absolute differences in rankings of agents' mates in two matchings. This class roughly includes our formulation with the caveat that we use attainable rankings and assigning each agent to their median attainable mate may not result in a matching.

[14] These points are demonstrated via the problem presented in Section 7.4.

[15] In Section 6.1, we clarify the connection between attainable mates and the rotation poset. Gusfield (1987) shows that a representation for the rotation poset can be constructed in $O(n^2)$ time.

[16] It follows from Irving, Leather & Gusfield (1987)[Theorem 5.2] that optimizing an additive function over the rotation poset can be computed in $O(n^2)$ time. Since our Lemma 2 in Section 6.1 shows that a modular assessment function can be represented as an additive function over the rotation poset, the conclusion follows.

5 Final comments

We conclude with a brief discussion of an important direction for future research. A major concern is the extension of our results to more general matching contexts that are rich with market-design applications. For such an extension, Blair (1984) provides a precursory result saying that for every distributive lattice, there is a marriage problem whose stable matchings lattice is order-isomorphic to the given distributive lattice. However, to explore this connection, it is critical that the set of stable matchings forms a distributive lattice. In this vein, Alkan (2001) and Alkan & Gale (2003) show that in the context of many-to-one and many-to-many matchings, the distributivity of the stable matchings lattice is guaranteed by strengthening the *substitutability* of the choice functions with *size monotonicity*. Under these restrictions Faenza & Zhang (2022) study algorithms for optimizing a modular function where preferences are replaced with choice functions, indicating that *modularity* preserves its appeal and relevance in these general domains. We believe that these findings pave the way for extending our results.

References

Afriat, S. N. (1967), 'The construction of utility functions from expenditure data', *International Economic Review* **8**(1), 67–77. 5

Alkan, A. (2001), 'On preferences over subsets and the lattice structure of stable matchings', *Review of Economic Design* **6**(1), 99–111. 20

Alkan, A. & Gale, D. (2003), 'Stable schedule matching under revealed preference', *Journal of Economic Theory* **112**(2), 289–306. 20

Balinski, M. & Sönmez, T. (1999), 'A tale of two mechanisms: Student placement', *Journal of Economic Theory* **84**, 73–94. 6

Birkhoff, G. (1937), 'Rings of sets', *Duke Mathematical Journal* **3**(3), 443–454. 25

Blair, C. (1984), 'Every finite distributive lattice is a set of stable matchings', *Journal of Combinatorial Theory, Series A* **37**(3), 353–356. 20

Can, B., Pourpouneh, M. & Storcken, T. (2023), 'Distance on matchings: An axiomatic approach', *Theoretical Economics* **18**(2), 597–632. 19

Chambers, C. P. & Echenique, F. (2008), 'Ordinal notions of submodularity', *Journal of Mathematical Economics* **44**(11), 1243–1245. 6

Chambers, C. P. & Echenique, F. (2009), 'Supermodularity and preferences', *Journal of Economic Theory* **144**(3), 1004–1014. 7

Chen, P., Egesdal, M., Pycia, M. & Yenmez, M. B. (2016), 'Median stable matchings in two-sided markets', *Games and Economic Behavior* **97**, 64–69. 9

Cheng, C. T. (2010), 'Understanding the generalized median stable matchings', *Algorithmica* **58**(1), 34–51. 9

Cheng, C. T., McDermid, E. & Suzuki, I. (2016), 'Eccentricity, center and radius computations on the cover graphs of distributive lattices with applications to stable matchings', *Discrete Applied Mathematics* **205**, 27–34. 5, 9

Echenique, F., Robinson-Cortes, A. & Yariv, L. (2022), 'An experimental study of decentralized matching', *mimeo* . 3, 5

Ehlers, L. & Klaus, B. (2006), 'Efficient priority rules', *Games and Economic Behavior* **55**(2), 372–384. 6

Faenza, Y. & Zhang, X. (2022), 'Affinely representable lattices, stable matchings, and choice functions', *Mathematical Programming* pp. 1–40. 20

Gale, D. & Shapley, L. S. (1962), 'College admissions and the stability of marriage', *American Mathematical Monthly* **69**, 9–15. 3

Gusfield, D. (1987), 'Three fast algorithms for four problems in stable marriage', *SIAM Journal on Computing* **16**(1), 111–128. 19

Gusfield, D. & Irving, R. W. (1989), 'The stable marriage problem: structure and algorithms'. 3, 8, 12

Henzinger, M., Noe, A., Schulz, C. & Strash, D. (2020), 'Finding all global minimum cuts in practice', *arXiv preprint arXiv:2002.06948* . 4

Irving, R. W. (1985), 'An efficient algorithm for the stable roommates problem', *Journal of Algorithms* **6**(4), 577–595. 14, 23

Irving, R. W. & Leather, P. (1986), 'The complexity of counting stable marriages', *SIAM Journal on Computing* **15**(3), 655–667. 5, 14, 25, 26

Irving, R. W., Leather, P. & Gusfield, D. (1987), 'An efficient algorithm for the optimal stable marriage', *Journal of the ACM (JACM)* **34**(3), 532–543. 4, 10, 19

Klaus, B. & Klijn, F. (2006), 'Procedurally fair and stable matching', *Economic Theory* **27**(2), 431–447. 6

Klaus, B. & Klijn, F. (2010), 'Smith and Rawls share a room: stability and medians', *Social Choice and Welfare* **35**(4), 647–667. 9

Knuth, D. E. (1976), 'Marriages stables', *Technical report* . 9

Kojima, F. & Manea, M. (2010), 'Axioms for deferred acceptance', *Econometrica* **78**, 633–653. 6

Kreps, D. M. (1979), 'A representation theorem for "preference for flexibility"', *Econometrica: Journal of the Econometric Society* pp. 565–577. 6

Manlove, D. (2013), *Algorithmics of matching under preferences*, Vol. 2, World Scientific. 8

Masarani, F. & Gokturk, S. S. (1989), 'On the existence of fair matching algorithms', *Theory and Decision* **26**(3), 305–322. 6

Milgrom, P. & Shannon, C. (1994), 'Monotone comparative statics', *Econometrica: Journal of the Econometric Society* pp. 157–180. 6

Picard, J.-C. (1976), 'Maximal closure of a graph and applications to combinatorial problems', *Management Science* **22**(11), 1268–1272. 4, 10

Picard, J.-C. & Queyranne, M. (1980), On the structure of all minimum cuts in a network and applications, *in* 'Combinatorial Optimization II', Springer, pp. 8–16. 4

Schwarz, M. & Yenmez, M. B. (2011), 'Median stable matching for markets with wages', *Journal of Economic Theory* **146**(2), 619. 9

Sethuraman, J., Teo, C.-P. & Qian, L. (2006), 'Many-to-one stable matching: geometry and fairness', *Mathematics of Operations Research* **31**(3), 581–596. 9

Teo, C.-P. & Sethuraman, J. (1998), 'The geometry of fractional stable matchings and its applications', *Mathematics of Operations Research* **23**(4), 874–891. 3, 9

6 Appendix A: Stepping stones

6.1 Rotations and some preliminary observations

For a given fixed problem $\succ \in \mathcal{P}$, *rotations*–first introduced by Irving (1985)–are the incremental changes that need to be made so that a stable matching μ can be transformed into another stable matching μ' such that $\mu \rhd_M \mu'$ and there is no other stable matching μ'' such that $\mu \rhd_M \mu'' \rhd_M \mu'$.

$$\rho^{11} = [(m_1, w_1), (m_2, w_2)]$$
$$\rho^{12} = [(m_3, w_3), (m_4, w_4)]$$
$$\rho^{13} = [(m_5, w_5), (m_6, w_6)]$$
$$\rho^2 = [(m_1, w_2), (m_4, w_3), (m_5, w_6), (m_2, w_1), (m_3, w_4), (m_6, w_5)]$$
$$\rho^3 = [(m_1, w_3), (m_2, w_4), (m_3, w_5), (m_4, w_6), (m_5, w_1), (m_6, w_2)]$$
$$\rho^4 = [(m_1, w_4), (m_2, w_5), (m_3, w_6), (m_4, w_1), (m_5, w_2), (m_6, w_3)]$$

M \ W	w_1	w_2	w_3	w_4	w_5	w_6
m_1	1 \ 6	2 \ 5	3 \ 4	4 \ 3	5 \ 1	6 \ 2
m_2	2 \ 5	1 \ 6	6 \ 2	3 \ 4	4 \ 3	5 \ 1
m_3	5 \ 1	6 \ 2	1 \ 6	2 \ 5	3 \ 4	4 \ 3
m_4	4 \ 3	5 \ 1	2 \ 5	1 \ 6	6 \ 2	3 \ 4
m_5	3 \ 4	4 \ 3	5 \ 1	6 \ 2	1 \ 6	2 \ 5
m_6	6 \ 2	3 \ 4	4 \ 3	5 \ 1	2 \ 5	1 \ 6

Figure 4: The problem in Section 4 and the associated rotations.

Let μ^M and μ^W denote the men-optimal and women-optimal stable matchings, and μ be a stable matching such that $\mu \neq \mu^W$. Then, $\mu(m) \neq \mu^W(m)$ for some man m. For each such man m, define his **successor woman** at μ, denoted by $s_\mu(m)$, as the $\succ_m$-best attainable woman w such that $\mu(m) \succ_m w$ (and $m \succ_w \mu(m)$). A **rotation** ρ **exposed in** μ is an (ordered) cyclic sequence of distinct man-woman pairs $\rho = [(m_1, w_1), (m_2, w_2) \ldots, (m_k, w_k)]$ such that $m_i w_i \in \mu$ and $s_\mu(m_i) = w_{i+1}$ for each $i \in \{1, \ldots, k\}$, where the addition in the subscripts is modulo k. To **eliminate** a rotation ρ exposed in a stable matching μ, each man m_i in ρ is matched to w_{i+1} while all the pairs that are not in ρ are kept the same. As a result, we obtain another stable matching, denoted by $\mu \circlearrowright \rho$, such that $\mu \rhd_M \mu \circlearrowright \rho$ and there is no other stable matching μ' with $\mu \rhd_M \mu' \rhd_M \mu \circlearrowright \rho$.

Let $\mathcal{R}$ denote the set of all rotations exposed in some stable matching. A rotation ρ **precedes** another one ρ', denoted by $\rho \to \rho'$, if in order to obtain a stable matching in which ρ' is exposed, ρ must be eliminated first. We assume that a rotation precedes itself. A rotation ρ **immediately precedes** another rotation ρ' if $\rho \to \rho'$ and there is no other rotation ρ'' such that $\rho \to \rho'' \to \rho'$. A distinct pair of rotations ρ and ρ' are **independent** if none of them precedes the other. The pair $\langle \mathcal{R}, \to \rangle$ is called the **rotation poset**.

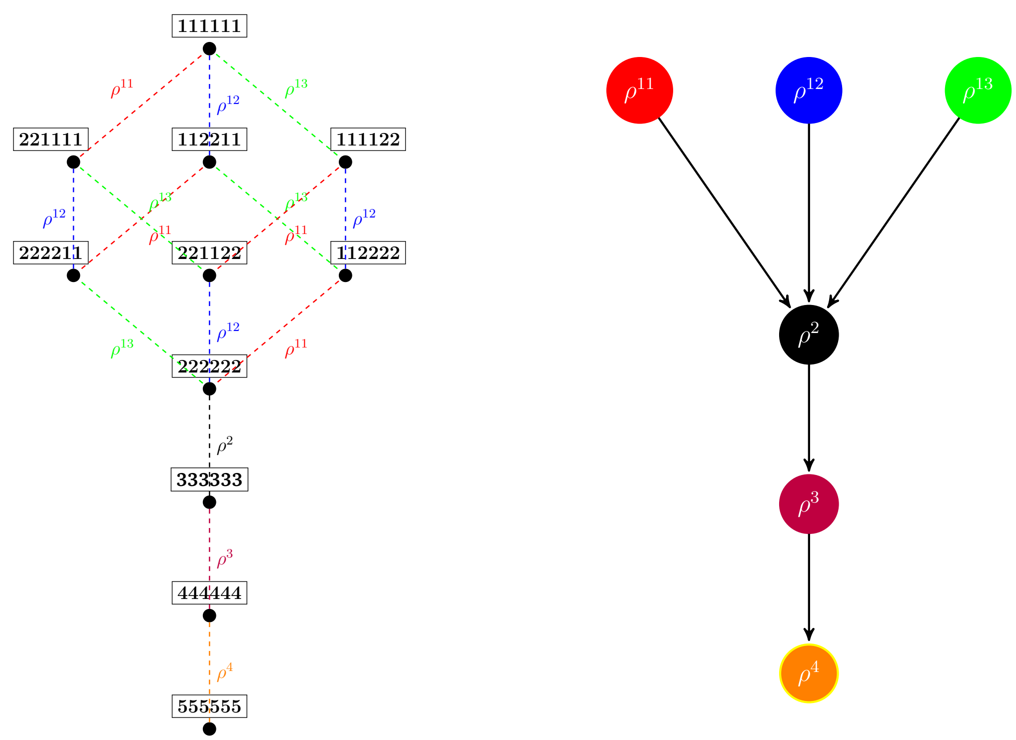


Figure 5: The stable matchings lattice and the rotation poset for the problem in Section 4 .

A subset of $\mathcal{R}$, generically denoted by R, is **closed** if whenever a rotation $\rho \in R$, then all the rotations that precede ρ are also in R. Let $Cl(\mathcal{R})$ denote the set of all closed subsets of $\mathcal{R}$. We suppress the reference to the specific problem $\succ \in \mathcal{P}$, whenever it is clear from the context. We will use the following basic properties of rotations. First, a man-woman pair (m, w) belongs to a rotation if and only if it appears in some stable matching and w is not the worst mate of m in all stable matchings; and a man-woman pair can be an element of at most one rotation (Irving & Leather 1986).

For each $\rho \in \mathcal{R}$, let N_ρ denote the set of agents who appear in rotation ρ. We note that if a pair of rotations ρ and ρ' are independent, then there is no agent who appears both in ρ and ρ', i.e. $N_\rho \cap N_{\rho'} = \emptyset$. Conversely, if there is no agent who appears both in ρ and ρ', then none of these rotations immediately precedes the other. In their main result, Irving & Leather (1986) show that the closed subsets of $\mathcal{R}$ endowed with the set containment relation $\langle Cl(\mathcal{R}), \subset \rangle$ is a lattice that is *order isomorphic*[17] to $\langle S, \rhd_M \rangle$. This result is parallel to Birkhoff's Representation Theorem (Birkhoff 1937) for distributive lattices.[18] Next, we make some simple observations using this result. Let $\succ \in \mathcal{P}$ be a given problem with the associated set of stable matching $\mathcal{S}$ and set of rotations $\mathcal{R}$. For each $\mu \in \mathcal{S}$, let R_μ be the associated closed subset of the rotation poset $\langle \mathcal{R}, \rightarrow \rangle$.

Lemma 1. *For each* $\mu \in \mathcal{S}, \sum_{mw \in \mu} (Rank_m^A(w) + Rank_w^A(m))$ *is the same.*

Proof. Let $\mu, \mu' \in \mathcal{S}$ such that $\mu' = \mu \circlearrowright \rho$ for some $\rho \in \mathcal{R}$, We first show that for each $(m, w) \in \rho$, we have

$$Rank_m^A(\mu'(m)) = Rank_m^A(w) + 1 \text{ and } Rank_w^A(\mu'(w)) = Rank_w^A(m) - 1 \tag{2}$$

To see this, note that since $\mu' = \mu \circlearrowright \rho$, for each $(m, w) \in \rho$, we have $w \succ_m \mu'(m)$ and $\mu'(w) \succ_w m$. Since there is also no other stable matching μ'' such that $\mu \rhd_M \mu'' \rhd_M \mu'$, it directly follows that (2) holds. It directly follows that $\sum_{mw \in \mu'} Rank_m^A(w) = \sum_{mw \in \mu} Rank_m^A(w) + |\rho|$ and $\sum_{mw \in \mu'} Rank_w^A(m) = \sum_{mw \in \mu} Rank_w^A(m) - |\rho|$. Since each $\mu \in \mathcal{S}$ can be obtained from μ^M by sequentially eliminating the rotations in $R_\mu \setminus R_{\mu^M}$, we reach the conclusion. □

[17] Given two posets $(S, \leq_S)$ and $(T, \leq_T)$, an *order isomorphism* from $(S, \leq_S)$ to $(T, \leq_T)$ is a bijective function f from S to T that is an *order embedding*, i.e. for each $x, y \in S$, $x \leq_S y$ if and only if $f(x) \leq_T f(y)$.

[18] Asserting that for a distributive lattice L, the closed subsets of the partially ordered set induced by its *join-irreducible elements* form a distributive lattice that is isomorphic to L.

Lemma 2. *A function $F : \mathcal{S} \to \mathbb{R}$ is modular if and only if there exists an additive set function $G : Cl(\mathcal{R}) \to \mathbb{R}$ such that for each $\mu \in \mathcal{S}$, $F(\mu) = G(R_\mu)$.*

Proof. Let $R \in Cl(\mathcal{R})$. Then, it follows from Irving & Leather (1986) that there exists $\mu \in \mathcal{S}$ such that $R = R_\mu$. We define $G : Cl(\mathcal{R}) \to \mathbb{R}$ such that for each $\mu \in \mathcal{S}$, $G(R_\mu) = F(\mu)$. Now, let $\mu, \mu' \in \mathcal{S}$. Since $\langle Cl(\mathcal{R}), \subset \rangle$ is order isomorphic to $\langle S, \rhd_M \rangle$, we have $R_{\mu \vee \mu'} = R_\mu \cap R_{\mu'}$ and $R_{\mu \wedge \mu'} = R_\mu \cup R_{\mu'}$. It follows that F is modular if and only if $G(R_\mu \cup R_{\mu'}) = G(R_\mu) + G(R_{\mu'}) - G(R_\mu \cap R_{\mu'})$. $\square$

Lemma 3. *Let $G : Cl(\mathcal{R}) \to \mathbb{R}$ be an additive set function. Then, there exists $g : \mathcal{R} \to \mathbb{R}$ such that for each $R \in Cl(\mathcal{R})$, $G(R) = G(\emptyset) + \sum_{\rho \in R} g(\rho)$.*

Proof. Let $\rho \in \mathcal{R}$ and define its closure $C(\rho) = \{\rho' \in \mathcal{R} \mid \rho' \to \rho\}$. Since the precedence relation is transitive, $C(\rho) \backslash \{\rho\}$ is closed. Therefore, for each $\rho \in \mathcal{R}$, we can define $g(\rho) = G(C(\rho)) - G(C(\rho) \setminus \{\rho\})$. Then, for each $\rho \in \mathcal{R}$, we have $G(C(\rho)) = G(\emptyset) + \sum_{\rho' \in C(\rho)} g(\rho')$. Now, let $R \in Cl(\mathcal{R})$. Then, since $R = \bigcup_{\rho \in R} C(\rho)$, by additivity of G, we conclude that $G(R) = G(\emptyset) + \sum_{\rho \in R} g(\rho)$. $\square$

Lemma 4. *Let $F : \mathcal{S} \to \mathbb{R}$ be a modular function. Then, for each $i \in N$, there exists $F_i : A_i \to \mathbb{R}$ such that for each $\mu \in \mathcal{S}$, $F(\mu) = \sum_{i \in N} F_i(\mu(i))$.*

Proof. It follows from Lemma 2 and Lemma 3 that there exists $g : \mathcal{R} \to \mathbb{R}$ such that for each $\mu \in \mathcal{S}$, $F(\mu) = G(\emptyset) + \sum_{\rho \in R_\mu} g(\rho)$ for some $G(\emptyset) \in \mathbb{R}$. Now, for each $\rho \in \mathcal{R}$ and $i \in N_\rho$, define $g_i(\rho) = g(\rho)/|N_\rho|$. Note that, by construction, we have $g(\rho) = \sum_{i \in N_\rho} g_i(\rho)$. Next, for each $i \in N$ and $j \in A_i$, if i and j are matched at the men-optimal stable matching, then define $F_i(j) = G(\emptyset)/2n$. Otherwise, let ρ_{ij} be the unique rotation elimination of which makes i matched to j, and define

$$F_i(j) = \sum_{\{\rho | \rho \to \rho_{ij}\}} g_i(\rho). \tag{3}$$

Now, let $\mu \in \mathcal{S}$. Since $R_\mu \in Cl(\mathcal{R})$, we have $R_\mu = \bigcup_{i \in N} \{\rho | \rho \to \rho_{i\mu(i)}\}$. It follows that

$$\sum_{\rho \in R_\mu} g(\rho) = \sum_{i \in N} \sum_{\{\rho | \rho \to \rho_{i\mu(i)}\}} g_i(\rho). \tag{4}$$

By substituting (3) into (4), we obtain that $F(\mu) = \sum_{i \in N} F_i(\mu(i))$. $\square$

6.2 An order isomorphism result

Let π be a stable matching rule and $\succ \in \mathcal{P}$ be a problem with the associated rotation poset $\langle \mathcal{R}, \rightarrow \rangle$ such that $\pi(\succ)$ is a sublattice of $\langle \mathcal{S}(\succ), \rhd_M \rangle$. First, let $\bar{\mu}(\underline{\mu})$ be the $\rhd_M$-best(worst) matching in $\pi(\succ)$. Then, define $\mathcal{R}^\delta = R_{\underline{\mu}} \setminus R_{\bar{\mu}}$ and

$$Cl^0(\mathcal{R}^\delta) = \{R_\mu \setminus R_{\bar{\mu}} \mid \mu \in \pi(\succ)\}.$$

Evidently, elements of $Cl^0(\mathcal{R}^\delta)$ are not closed according to the precedence relation $\rightarrow$, unless $\bar{\mu}$ is the men-optimal stable matching at $\succ$. Next, we recursively define two set collections $\{X^i\}_{i=1}^K$ and $\{\Lambda^i\}_{i=1}^K$. Then, we prove a related structural result, Proposition 3, which will be an important stepping stone in proving Theorem 1.

For $k = 1$: Consider $min(Cl^0(\mathcal{R}^\delta), \subset)$ that consists of $R \in Cl^0(\mathcal{R}^\delta)$ such that there is no $R' \in Cl^0(\mathcal{R}^\delta) \setminus \emptyset$ with $R' \subsetneq R$. Let $X^1 = Cl^0(\mathcal{R}^\delta)$ and $\Lambda^1 = min(X^1, \subset)$.

For $k \geq 2$: Define $X^2 = \{R \setminus \bigcup_{R' \in \Lambda^1} R' \mid R \in X^1\}$ and $\Lambda^2 = min(X^2, \subset)$.[19] Similarly, for each $k \geq 1$, define $X^{k+1} = \{R \setminus \bigcup_{R' \in \Lambda^k} R' \mid R \in X^k\}$ and $\Lambda^{k+1} = min(X^{k+1}, \subset)$.

Let $K \geq 1$ be the smallest number such that $X^{K+1} = \emptyset$. Then, $\{X^k\}_{k=1}^K$ and $\{\Lambda^k\}_{k=1}^K$ are the ordered collection of the disjoint nonempty sets that are constructed. Define $\Lambda = \bigcup_{k=1}^K \Lambda^k$. We call each member of Λ a **hyper-rotation** and generically denote by λ. Figure 6 presents a demonstration of how Λ is formed.

Lemma 5. *For each $k \in \{1, \ldots, K\}$, $\langle X^k, \subset \rangle$ is a lattice.*

Proof. By induction, first, consider the case that $k = 1$, where $X^1 = Cl^0(\mathcal{R}^\delta)$. Since $\pi(\succ)$ is a sublattice of $\langle \mathcal{S}(\succ), \rhd_M \rangle$, $Cl^0(\mathcal{R}^\delta)$ is a sublattice of $\langle Cl(\mathcal{R}^\delta), \subset \rangle$. Therefore, $\langle X^1, \subset \rangle$ is a lattice. Next, for each $k \in \{1, \ldots, K-1\}$ assume that $\langle X^k, \subset \rangle$ is a lattice, and let $Q, Q' \in X^{k+1}$. By construction of X^{k+1}, there exist $R \in X^k$ and $R' \in X^k$ such that $R = Q \cup \bigcup_{\lambda \in \Lambda^k} \lambda$ and $R' = Q' \cup \bigcup_{\lambda \in \Lambda^k} \lambda$. Since $\langle X^k, \subset \rangle$ is a lattice, $R \cap R' \in X^k$ and $R \cup R' \in X^k$. Therefore, $Q \cap Q' \in X^{k+1}$ and $Q \cup Q' \in X^{k+1}$. Thus, we conclude that $\langle X^{k+1}, \subset \rangle$ is a lattice. □

[19]The elements of X^2 and Λ^2 are sets of rotations that are closed neither according to $\rightarrow$ nor according to $\rightrightarrows$ that is to be defined later.

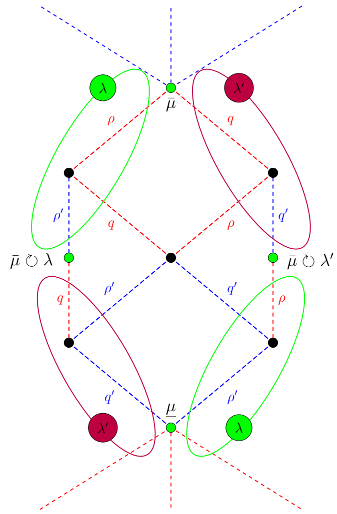


Figure 6: A demonstration of our construction for Λ, where $\pi(\succ)$ are the green labelled nodes.

Lemma 6. *Let $\rho \in \mathcal{R}^\delta$. Then, there exists unique $\lambda_\rho \in \Lambda$ such that $\rho \in \lambda_\rho$.*

Proof. First, note that by construction, $\{\Lambda^k\}_{k=1}^K$ is a collection of disjoint sets such that for each $\rho \in \mathcal{R}^\delta$, there exists $k \in \{1, \ldots, K\}$ and $\lambda \in \Lambda^k$ such that $\rho \in \lambda$. Next, let $k \in \{1, \ldots, K\}$ and $\lambda, \lambda' \in \Lambda^k$ be distinct. Since, by Lemma 5, $\langle X^k, \subset \rangle$ is a lattice, $\lambda \cap \lambda' \in X^k$. Then, $\Lambda^k = min(X^k, \subset)$ implies that $\lambda \cap \lambda' = \emptyset$. $\square$

Lemma 7. *Let $R \in Cl^0(\mathcal{R}^\delta)$. Then, $\{\lambda_\rho\}_{\rho \in R}$ partitions R.*

Proof. For each $k \in \{1, \ldots, K\}$, define $\Lambda_R^k = \{\lambda \in \Lambda^k \mid \lambda \subset R\}$. Recall that $X^1 = Cl^0(\mathcal{R}^\delta)$ and since $X^{K+1} = \emptyset$, we have $X^K = \Lambda^K$. Then, since $R \in Cl^0(\mathcal{R}^\delta)$, by the construction of $\{X^k\}_{k=1}^K$ and $\{\Lambda^k\}_{k=1}^K$, we have $R = \bigcup_{k=1}^K \bigcup_{\{\lambda \in \Lambda_R^k\}} \lambda$. Since for each $\rho \in R$, by Lemma 6, λ_ρ is the unique member of Λ with $\rho \in \lambda_\rho$, it follows that $R = \bigcup_{\{\rho \in R\}} \lambda_\rho$ and for each $\rho, \rho' \in R$, either $\lambda_\rho = \lambda_{\rho'}$ or $\lambda_\rho \cap \lambda_{\rho'} = \emptyset$. $\square$

Now, we are ready to prove Proposition 3, which we will use to prove Theorem 1. To introduce this result, first, recall that $\Lambda = \bigcup_{k=1}^K \Lambda^k$. Then, for each distinct $\lambda, \lambda' \in \Lambda$, λ **precedes** λ', denoted by $\lambda \rightrightarrows \lambda'$, if for each $R \in Cl^0(\mathcal{R}^\delta)$, $\lambda' \subset R$ implies that $\lambda \subset R$. Put differently, $\lambda \rightrightarrows \lambda'$ if

$$\lambda \subset \bigcap_{\{R \in Cl^0(\mathcal{R}^\delta) \mid \lambda' \subset R\}} R.$$

Note that $\langle \Lambda, \rightrightarrows \rangle$ is a finite poset. Let $Cl(\Lambda)$ be the closed subsets of Λ with respect to the precedence relation $\rightrightarrows$. In Proposition 3, we show that $\langle Cl^0(\mathcal{R}^\delta), \subset \rangle$ is *order isomorphic* to $\langle Cl(\Lambda), \subset \rangle$. To see this, define $\Lambda : Cl^0(\mathcal{R}^\delta) \to Cl(\Lambda)$ such that for each $R \in Cl^0(\mathcal{R}^\delta)$, $\Lambda(R) = \{\lambda_\rho\}_{\rho \in R}$.

Proposition 3. *The Λ mapping induces an order isomorphism between $\langle Cl^0(\mathcal{R}^\delta), \subset \rangle$ and $\langle Cl(\Lambda), \subset \rangle$.*

Proof. Let $R \in Cl^0(\mathcal{R}^\delta)$. First, we verify that $\Lambda(R) \in Cl(\Lambda)$. To see this, let $\lambda \in \Lambda(R)$ and $\lambda' \in \Lambda$ such that $\lambda' \rightrightarrows \lambda$. Then, since $R \in Cl^0(\mathcal{R}^\delta)$ with $\lambda \subset R$, it follows from $\lambda' \rightrightarrows \lambda$ that $\lambda' \subset R$. Therefore, $\lambda' \in \Lambda(R)$. To see that Λ is one-to-one, note that for each $R, R' \in Cl^0(\mathcal{R}^\delta)$, if $R \neq R'$ then $\Lambda(R) \neq \Lambda(R')$. To see that Λ is an order embedding, by Lemma 7, for each $R, R' \in Cl^0(\mathcal{R}^\delta)$, $\{\lambda_\rho\}_{\rho \in R}$ partitions R and $\{\lambda_{\rho'}\}_{\rho' \in R'}$ partitions R'. It follows that $R \subset R'$ if and only if $\{\lambda_\rho\}_{\rho \in R} \subset \{\lambda_\rho\}_{\rho \in R'}$.

Finally, we show that Λ is onto. To see this, let $Q \in Cl(\Lambda)$ and define $R_Q = \bigcup_{\lambda \in Q} \lambda$. We show that $R_Q \in Cl^0(\mathcal{R}^\delta)$ and $\Lambda(R_Q) = Q$. To see this, let $\lambda \in \Lambda$ and recall that, by construction, there exists $R \in Cl^0(\mathcal{R}^\delta)$ such that $\lambda \subset R$. Therefore, we can define $R_\lambda = \bigcap_{\{R \in Cl^0(\mathcal{R}^\delta) | \lambda \subset R\}} R$. Then, since $Cl^0(\mathcal{R}^\delta)$ is a lattice, we have $R_\lambda \in Cl^0(\mathcal{R}^\delta)$ and $\bigcup_{\lambda \in Q} R_\lambda \in Cl^0(\mathcal{R}^\delta)$. Next, to conclude, we show that $R_Q = \bigcup_{\lambda \in Q} R_\lambda$.

Since for each $\lambda \in Q$, $\lambda \subset R_\lambda$ and $R_Q = \bigcup_{\lambda \in Q} \lambda$, we have $R_Q \subset \bigcup_{\lambda \in Q} R_\lambda$. To see the converse, let $\lambda \in Q$ and $\rho \in R_\lambda$. We show that $\rho \in R_Q$. Since $R_\lambda \in Cl^0(\mathcal{R}^\delta)$, by Lemma 7, $\{\lambda_{\rho'}\}_{\{\rho' \in R_\lambda\}}$ partitions R_λ. Therefore, $\lambda_\rho \subset R_\lambda$. Then, it follows from the construction of R_λ that for each $R \in Cl^0(\mathcal{R}^\delta)$, if $\lambda \subset R$ then $\lambda_\rho \subset R$, that is $\lambda_\rho \rightrightarrows \lambda$. Now, since $Q \in Cl(\Lambda)$ and $\lambda \in Q$, $\lambda_\rho \rightrightarrows \lambda$ implies that $\lambda_\rho \in Q$. Since $R_Q = \bigcup_{\lambda \in Q} \lambda$, we have $\lambda_\rho \subset R_Q$ indicating that $\rho \in R_Q$. Thus, we conclude that $R_Q = \bigcup_{\lambda \in Q} R_\lambda$. Then, it directly follows from the formulation of Λ that $\Lambda(R_Q) = Q$. $\square$

7 Appendix B

7.1 Proof of Theorem 1

Only if part: Let π be a modular stable matching rule. It directly follows from modularity that $\pi(\succ)$ is a sublattice of $\langle \mathcal{S}(\succ), \rhd_M \rangle$. Moreover, by Lemma 2, there exists an additive set function $G : Cl(\mathcal{R}) \to \mathbb{R}$ such that $\pi(\succ) = \text{argmin}_{\mu \in \mathcal{S}(\succ)} G(R_\mu)$ and by Lemma 3 there exists $g : \mathcal{R} \to \mathbb{R}$ such that for each $R \in Cl(\mathcal{R})$, $G(R) = G(\emptyset) + \sum_{\rho \in R} g(\rho)$.

To see that π satisfies *convexity*, let $\mu^*, \mu^{**} \in \pi(\succ)$ and $\mu \in \mathcal{S}(\succ)$ such that $\mu(m) \in \{\mu^*(m), \mu^{**}(m)\}$ for each $m \in M$. We show that $\mu \in \pi(\succ)$. First, let $\mu' = \mu^* \vee \mu^{**}$ and $\mu'' = \mu^* \wedge \mu^{**}$. Note that, we have $\mu(m) \in \{\mu'(m), \mu''(m)\}$ for each $m \in M$, and $\mu', \mu'' \in \pi(\succ)$ since $\pi(\succ)$ is a sublattice. Therefore, if $\mu = \mu'$ or $\mu = \mu''$ then we conclude that $\mu \in \pi(\succ)$; if not then $\mu' \rhd_M \mu \rhd_M \mu''$, which implies that $R_{\mu'} \subsetneq R_\mu \subsetneq R_{\mu''}$. Let $P = R_\mu \setminus R_{\mu'}$ and $P' = R_{\mu''} \setminus R_\mu$.

Next, we show that each $\rho \in P$ and each $\rho' \in P'$ are independent. To see this, recall that N_ρ denotes the set of agents in a rotation ρ. Let $\rho \in P$ and $i \in N_\rho$, then in moving from μ' to μ it must be that agent i's mate is changed and $\mu(i) = \mu''(i)$, and in moving from μ to μ'', i's mate can not change, that is there is no $\rho' \in P'$ such that $i \in N_{\rho'}$. Therefore, for each $\rho \in P$ and $\rho' \in P'$, $N_\rho \cap N_{\rho'} = \emptyset$, and there is no $\rho \in P$ that immediately precedes any $\rho' \in P'$. It follows that each $\rho \in P$ and each $\rho' \in P'$ are independent.

Now, since each $\rho \in P$ and each $\rho' \in P'$ are independent, $R_{\mu'} \cup P' \in Cl(\mathcal{R})$. Let $\mu''' \in \mathcal{S}(\succ)$ be such that $R_{\mu'''} = R_{\mu'} \cup P'$. Finally, to get a contradiction that μ' minimizes G, we show that $G(R_{\mu'''}) < G(R_{\mu'})$. Since $\mu', \mu'' \in \pi(\succ)$, we have $G(R_{\mu'}) = G(R_{\mu''})$. Since G is additive and $\{P, P'\}$ partitions $R_{\mu''} \setminus R_{\mu'}$, it follows that $\sum_{\rho \in P} g(\rho) + \sum_{\rho' \in P'} g(\rho') = 0$. Now, if $\mu \notin \pi(\succ)$, then we must have $\sum_{\rho \in P} g(\rho) > 0$, which implies that $\sum_{\rho' \in P'} g(\rho') < 0$. Then, we have $G(R_{\mu'''}) < G(R_{\mu'})$. Thus, we conclude that $\mu \in \pi(\succ)$.

If part: Let π be a stable matching rule that satisfies *convexity*. Let $\succ \in \mathcal{P}$ be a problem with the associated rotation poset $\langle \mathcal{R}, \to \rangle$. To show that there exists a modular fairness measure

$F : \mathcal{S}(\succ) \to \mathbb{Z}$ such that $\pi(\succ) = \mathrm{argmin}_{\mu \in \mathcal{S}(\succ)} F(\mu)$, by Lemma 2, it is sufficient to show that there exists an additive set function $G : Cl(\mathcal{R}) \to \mathbb{R}$ such that $\pi(\succ) = arg_{\mu \in \mathcal{S}(\succ)} min\ G(R_\mu)$.

It directly follows from *convexity* of π that $\pi(\succ)$ is a sublattice of $\langle \mathcal{S}(\succ), \rhd_M \rangle$. Therefore, Proposition 3 holds for π. In what follows, we assume that the formal objects, such as $Cl^0(\mathcal{R}^\delta)$ and Λ, defined in Section 6.2 are associated with π. Note that for each $\mu \in \mathcal{S}(\succ)$, $\mu \in \pi(\succ)$ if and only if $R_\mu \setminus R_{\bar{\mu}} \in Cl^0(\mathcal{R}^\delta)$. Therefore, to prove the result, we show that there exists an additive set function $G : Cl(\mathcal{R}) \to \mathbb{R}$ such that for each $R \in Cl(\mathcal{R})$, R minimizes G if and only if $R \setminus R_{\bar{\mu}} \in Cl^0(\mathcal{R}^\delta)$. Next, by using Proposition 3 and *convexity*, we prove an important structural result that paves the way for constructing the desired additive set function.

Lemma 8 (Partition lemma). *Let $\lambda \in \Lambda$ that contains at least two rotations and $\{P, P'\}$ be a partition of λ. Then, there exist $\rho \in P$ and $\rho' \in P'$ such that $\rho \to \rho'$ or $\rho' \to \rho$.*

Proof. By contradiction, suppose that each $\rho \in P$ and $\rho' \in P'$ are independent. Let $\lambda \in \Lambda^j$ for some $j \in \{1, \ldots, K\}$. Recall that $\Lambda^j = min(X^j, \subset)$, and consider the set $A = \bigcup_{k=1}^{j-1} \Lambda^k$, in the case that $j = 1$, assume that $A = \emptyset$. Then, we have $A \in Cl(\Lambda)$. By construction of $\langle \Lambda, \rightrightarrows \rangle$, for each $\lambda' \in \Lambda$, if $\lambda' \rightrightarrows \lambda$, then $\lambda' \in \Lambda^i$ for some $i < j$. Therefore, we have $A \cup \{\lambda\} \in Cl(\Lambda)$. Then, by Proposition 3, there exist $R, R' \in Cl^0(\mathcal{R}^\delta)$ such that $\Lambda(R) = A$ and $\Lambda(R') = A \cup \{\lambda\}$. Let μ and μ' be the stable matchings in $\mathcal{S}(\succ)$ associated with R and R', i.e. $R = R_\mu \setminus R_{\bar{\mu}}$ and $R' = R_{\mu'} \setminus R_{\bar{\mu}}$.

First, we show that there is no $\mu'' \in \pi(\succ)$ with $\mu \rhd_M \mu'' \rhd_M \mu'$. Otherwise, let $R'' = R_{\mu''} \setminus R_{\bar{\mu}}$. Since $\mu'' \in \pi(\succ)$, we have $R'' \in Cl^0(\mathcal{R}^\delta)$ and since $\mu \rhd_M \mu'' \rhd_M \mu'$, we have $R \subsetneq R'' \subsetneq R'$. Then, by Proposition 3, $\Lambda(R'') \in Cl(\Lambda)$ and $\Lambda(R) \subsetneq \Lambda(R'') \subsetneq \Lambda(R')$. But, since $\Lambda(R) = A$ and $\Lambda(R') = A \cup \{\lambda\}$, there can not exist such $\Lambda(R'')$, a contradiction.

Now, let μ'' be the matching obtained from μ by eliminating all the rotations in P. Since each $\rho \in P$ and $\rho' \in P'$ are independent, $N_\rho \cap N_{\rho'} = \emptyset$. Therefore, in moving from μ to μ'' it must be that if an agent i's mate is changed, then $\mu''(i) = \mu'(i)$. It follows that for each $m \in M$, $\mu''(m) \in \{\mu(m), \mu'(m)\}$. Then, by *convexity*, we have $\mu'' \in \pi(\succ)$ contradicting there is no $\mu'' \in \pi(\succ)$ with $\mu \rhd_M \mu'' \rhd_M \mu'$. $\square$

We are now ready to construct an additive function $H : Cl(\mathcal{R}^\delta) \to \mathbb{Z}$ such that for each $R \in Cl(\mathcal{R}^\delta)$, $H(R) = 0$ if and only if $R \in Cl^0(\mathcal{R}^\delta)$. Let $\lambda \in \Lambda$ and define $\lambda^\downarrow$ as the set of rotations in λ that has no successor in λ, i.e. $\lambda^\downarrow = \{q \in \lambda \mid \text{there is no } \rho' \in \lambda \text{ with } q \to \rho'\}$ (we denote a generic element of $\lambda^\downarrow$ by q). Similarly define $\lambda^\uparrow$ as the set of rotations in λ that has no predecessor in λ, i.e. $\lambda^\uparrow = \{\rho \in \lambda \mid \text{there is no } \rho' \in \lambda \text{ with } \rho' \to \rho\}$. Since precedence is a transitive relation and λ is a finite set, $\lambda^\downarrow \neq \emptyset$ and $\lambda^\uparrow \neq \emptyset$. If λ is not a singleton, then $\lambda^\uparrow \cap \lambda^\downarrow = \emptyset$. Otherwise, suppose that there exists $\rho \in \lambda^\uparrow \cap \lambda^\downarrow$ and consider $\{p\}$ and $\lambda \setminus \{p\}$, which partitions λ. Then, there is no $\rho' \in \lambda$ such that $\rho \to \rho'$ (since $\rho \in \lambda^\downarrow$) or $\rho' \to \rho$ (since $\rho \in \lambda^\uparrow$). But, this contradicts to Lemma 8. Thus, we conclude that $\lambda^\uparrow \cap \lambda^\downarrow = \emptyset$ whenever λ is not a singleton. Now, let $\lambda^{\downdownarrows}$ and $\lambda^{\upuparrows}$ be any pair of set of rotations in λ that respectively contains $\lambda^\downarrow$ and $\lambda^\uparrow$ such that $\lambda^{\upuparrows} \cap \lambda^{\downdownarrows} = \emptyset$.

Remark 2. *For the current proof, we can assume that $\lambda^{\upuparrows} = \lambda^\uparrow$ and $\lambda^{\downdownarrows} = \lambda^\downarrow$. We present this general construction foreseeing that it will crucial in proving Theorem 2 and similar results.*

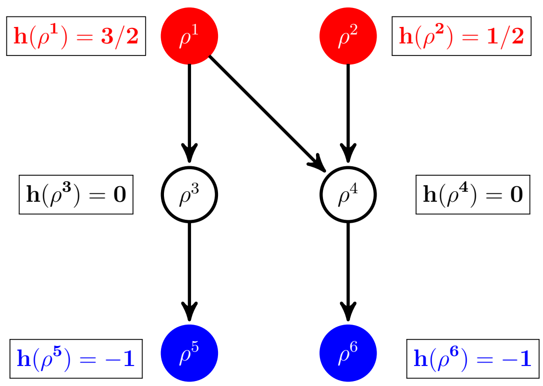
Let $\lambda \in \Lambda$ that is not a singleton and $\rho \in \lambda$. Then, define $\lambda^{\downdownarrows}(\rho)$ as the set of rotations in $\lambda^{\downdownarrows}$ that are preceded by ρ, i.e. $\lambda^{\downdownarrows}(\rho) = \{q \in \lambda^{\downdownarrows} \mid \rho \to q\}$. Similarly, define $\lambda^{\upuparrows}(\rho)$ as the set of rotations in $\lambda^{\upuparrows}$ that precede ρ, i.e. $\lambda^{\upuparrows}(\rho) = \{\rho' \in \lambda^{\upuparrows} \mid \rho' \to \rho\}$.

Next, recall that, by Lemma 6, for each $\rho \in \mathcal{R}^\delta$, there exists unique $\lambda_\rho \in \Lambda$ such that $\rho \in \lambda_\rho$. Therefore, we can define $h : \mathcal{R}^\delta \to \mathbb{Z}$ such that for each $\rho \in \mathcal{R}^\delta$, if $\lambda_\rho = \{\rho\}$, then $h(\rho) = 0$; if not, then let $\lambda = \lambda_\rho$ and define

$$h(\rho) = \begin{cases} -1 & \text{if } \rho \in \lambda^{\downdownarrows}, \\ \displaystyle\sum_{q \in \lambda^{\downdownarrows}(\rho)} \frac{1}{|\lambda^{\upuparrows}(q)|} & \text{if } \rho \in \lambda^{\upuparrows}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Figure 7 demonstrates this construction. For each nonempty $Q \subset \mathcal{R}^\delta$, define $H(Q) = \sum_{\rho \in Q} h(\rho)$ and $H(\emptyset) = 0$. Note that by the construction of h, for each $\lambda \in \Lambda$, $H(\lambda) = 0$. The construction of h together with Lemma 8 guarantees that the following assertion holds.

Lemma 9. *Let $\lambda \in \Lambda$ and $Q \subset \lambda^{\downdownarrows}$ be a nonempty set of rotations such that $\lambda^\downarrow \setminus Q \neq \emptyset$. Then, we have $H(\bigcup_{q \in Q} \lambda^{\upuparrows}(q)) > |Q|$.*



	$\lambda^{\downarrow}(\rho)$	$\lambda^{\uparrow}(\rho)$
ρ_1	$\{\rho_5, \rho_6\}$	$\{\rho_1\}$
ρ_2	$\{\rho_6\}$	$\{\rho_2\}$
ρ_3	$\{\rho_5\}$	$\{\rho_1\}$
ρ_4	$\{\rho_6\}$	$\{\rho_1, \rho_2\}$
ρ_5	$\{\rho_5\}$	$\{\rho_1\}$
ρ_6	$\{\rho_6\}$	$\{\rho_1, \rho_2\}$

Figure 7: Values of the h function, where $\lambda^{\Uparrow} = \lambda^{\uparrow} = \{\rho^1, \rho^2\}$ and $\lambda^{\Downarrow} = \lambda^{\downarrow} = \{\rho^5, \rho^6\}$.

Proof. To see this, first, let $\lambda^{\Uparrow}(Q) = \bigcup_{q \in Q} \lambda^{\Uparrow}(q)$ and note that, by the construction of h, $H(\lambda^{\Uparrow}(Q)) \geq |Q|$. In what follows, we show that there exists $\rho^* \in \lambda^{\Uparrow}(Q) \cap \lambda^{\uparrow}$ such that $\lambda^{\Downarrow}(\rho^*)$ contains some $q' \in \lambda^{\downarrow} \setminus Q$. Thus, we will conclude that $h(\rho^*) > \sum_{q \in Q \cap \lambda^{\Downarrow}(\rho^*)} \frac{1}{|\lambda^{\Uparrow}(q)|}$ and $H(\lambda^{\Uparrow}(Q)) > |Q|$. To see this, let P be the set of rotations in λ that precede an element of Q. Note that $P \neq \lambda$, since $\lambda^{\downarrow} \setminus Q \neq \emptyset$. Then, consider P and $\lambda \setminus P$, which partitions λ. It follows from Lemma 8 that there exists $\rho \in P$ and $\rho' \in \lambda \setminus P$ such that $\rho \to \rho'$ or $\rho' \to \rho$. By our choice of P, the latter is not possible, so we have $\rho \to \rho'$. Since $\rho' \in \lambda \setminus P$, there exists $q' \in \lambda^{\downarrow} \setminus Q$ such that $\rho' \to q'$, and thus $\rho \to q'$. Since $\rho \in P$, there exists $\rho^* \in \lambda^{\Uparrow}(Q) \cap \lambda^{\uparrow}$ such that $\rho^* \to \rho$. Thus, we conclude that $\rho^* \to q'$ as desired. $\square$

Next, we show that by restricting the domain of H to $Cl(\mathcal{R}^\delta)$, we will obtain the desired additive function. Figure 8 presents a demonstration of our construction.

Lemma 10. *For each $R \in Cl(\mathcal{R}^\delta)$, $H(R) = 0$ if and only if $R \in Cl^0(\mathcal{R}^\delta)$.*

Proof. (*If part*) As we noted before, by the construction of h, for each $\lambda \in \Lambda$, we have $H(\lambda) = 0$. Since, by Lemma 7, for each $R \in Cl^0(\mathcal{R}^\delta)$, $\{\lambda_\rho\}_{\rho \in R}$ partitions R, it follows that $H(R) = 0$.

(*Only if part*) Let $R \in Cl(\mathcal{R}^\delta)$ such that $R \notin Cl^0(\mathcal{R}^\delta)$. We show that $H(R) > 0$. First, recall that, by Lemma 6, for each $\rho \in \mathcal{R}^\delta$, there exists unique $\lambda \in \Lambda$ such that $\rho \in \lambda$. Therefore, $\{R \cap \lambda\}_{\lambda \in \Lambda}$ partitions R. Since H is additive, it follows that $H(R) = \sum_{\lambda \in \Lambda} H(R \cap \lambda)$.

Since $R \in Cl(\mathcal{R}^\delta)$ but $R \notin Cl^0(\mathcal{R}^\delta)$, by Proposition 3, there exists $\lambda \in \Lambda$, such that $R \cap \lambda \neq \emptyset$ and $\lambda \setminus R \neq \emptyset$. Therefore, λ is not a singleton, and by the construction of h, for each $\rho \in \lambda$, $h(\rho) <$

0 ($h(\rho) > 0$) if and only if $\rho \in \lambda^{\downdownarrows}$ ($\rho \in \lambda^{\upuparrows}$). Thus, we have $H(R \cap \lambda) = H(R \cap \lambda^{\upuparrows}) + H(R \cap \lambda^{\downdownarrows})$.

Since R is closed and for each $\rho \in \lambda$, either $\rho \in \lambda^{\upuparrows}$ or ρ is preceded by a rotation in $\lambda^{\uparrow} \subset \lambda^{\upuparrows}$, it follows from $R \cap \lambda \neq \emptyset$ that $R \cap \lambda^{\upuparrows} \neq \emptyset$. Now, first, suppose that $R \cap \lambda^{\downdownarrows} = \emptyset$. Then, $H(R \cap \lambda^{\downdownarrows}) = 0$, and thus $H(R \cap \lambda) > 0$. Next, suppose that $R \cap \lambda^{\downdownarrows} \neq \emptyset$ and let $Q = R \cap \lambda^{\downdownarrows}$. Note that, since R is closed, we have $\bigcup_{q \in Q} \lambda^{\upuparrows}(q) = R \cap \lambda^{\upuparrows}$. Therefore, $H(R \cap \lambda) = H(\bigcup_{q \in Q} \lambda^{\upuparrows}(q)) - |Q|$.

Finally, by Lemma 9, we will conclude that $H(R \cap \lambda) > 0$. To apply the lemma, we need to show that $\lambda^{\downarrow} \setminus Q \neq \emptyset$. But, if we would have $\lambda^{\downarrow} \subset Q$, then since R is closed and for each $\rho \in \lambda$, either $\rho \in \lambda^{\downarrow}$ or ρ precedes a rotation in $\lambda^{\downarrow}$, we must have $\lambda \subset R$, contradicting that $\lambda \setminus R \neq \emptyset$. $\square$

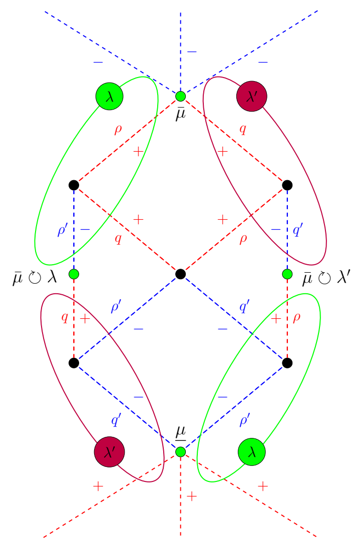


Figure 8: A demonstration of our construction of h and g.

Next, we extend the domain of H from $Cl(\mathcal{R}^{\delta})$ to $Cl(\mathcal{R})$. For this, first define $g : \mathcal{R} \to \mathbb{Z}$ such that for each $\rho \in \mathcal{R}$,

$$g(\rho) = \begin{cases} -1 & \text{if } \rho \in R_{\bar{\mu}}, \\ 1 & \text{if } \rho \notin R_{\underline{\mu}}, \\ h(\rho) & \text{otherwise, i.e. } \rho \in \mathcal{R}^{\delta}. \end{cases}$$

Then, define the additive function $G : Cl(\mathcal{R}) \to \mathbb{R}$ such that $G(\emptyset) = 0$ and for each nonempty $R \in Cl(\mathcal{R})$, $G(R) = \sum_{\rho \in R} g(\rho)$.

It follows from the construction of g that for each $R \in Cl(\mathcal{R})$, if R minimizes G, then $R_{\bar{\mu}} \subset R \subset R_{\underline{\mu}}$. Since $R \in Cl(\mathcal{R})$, this means that $R \setminus R_{\bar{\mu}} \in Cl(\mathcal{R}^{\delta})$. Then, since for each $\rho \in \mathcal{R}^{\delta}$, $g(\rho) = h(\rho)$, it directly follows from Lemma 10 that for each $R \in Cl(\mathcal{R})$, R minimizes G if and only if $R \setminus R_{\bar{\mu}} \in Cl^0(\mathcal{R}^{\delta})$. Thus, we complete the proof.

7.2 Proof of Proposition 2

The *if part* was proven in the main text. For the *only if part*, let $\succ \in \mathcal{P}$. Since π satisfies *convexity*, $\pi(\succ)$ is a sublattice of $\langle \mathcal{S}(\succ), \rhd_M \rangle$. To see that π satisfies *IIR*, let $\bar{\mu}(\underline{\mu})$ be the $\rhd_M$-best(worst) matching in $\pi(\succ)$. Recall that we obtain the men(women)-optimal stable matching for the problem $\succ^{\pi}$ by running the men(women)-proposing Gale-Shapley algorithm, in which each man(woman) proposes to women(men) by following his(her) transformed preference list. It follows that the men(women)-optimal stable matching for the problem $\succ^{\pi}$ is $\bar{\mu}(\underline{\mu})$. Therefore, for each $\mu \in \mathcal{S}(\succ)$, if $\mu \rhd_M \bar{\mu}$ or $\underline{\mu} \rhd_M \mu$, then $\mu \notin \mathcal{S}(\succ^{\pi})$. Now, let $\mu \in \mathcal{S}(\succ)$ be such that $\bar{\mu} \rhd_M \mu \rhd_M \underline{\mu}$. We show that if $\mu \in \mathcal{S}(\succ^{\pi})$, then $\mu \in \pi(\succ)$. By contradiction, suppose that $\mu \in \mathcal{S}(\succ^{\pi}) \setminus \pi(\succ)$. Assume without loss of generality that there is no other $\hat{\mu} \in \mathcal{S}(\succ^{\pi}) \setminus \pi(\succ)$ such that $\hat{\mu} \rhd_M \mu$.

We show that there exists $m \in M$ such that $\mu(m) \notin \pi_m(\succ)$. To see this, let $\mu', \mu'' \in \pi(\succ)$ be the $\rhd_M$-worst and $\rhd_M$-best matchings such that $\mu' \rhd_M \mu \rhd_M \mu''$. Since π satisfies *convexity* and $\mu \notin \pi(\succ)$, there exists $m \in M$ such that $\mu(m) \notin \{\mu'(m), \mu''(m)\}$. Next, we show that $\mu(m) \notin \pi_m(\succ)$. By contradiction, suppose that there exists $\tilde{\mu} \in \pi(\succ)$ with $\tilde{\mu}(m) = \mu(m)$. We show that this contradicts that there is no other $\hat{\mu} \in \mathcal{S}(\succ^{\pi}) \setminus \pi(\succ)$ such that $\hat{\mu} \rhd_M \mu$. To see this, let $\mu^* = (\mu' \wedge \tilde{\mu}) \vee \mu''$. Since $\pi(\succ)$ is a lattice, $\mu^* \in \pi(\succ)$ and therefore $\mu^* \neq \mu$. Moreover, $\mu^* \neq \mu'$ and $\mu^* \neq \mu''$, since $\mu^*(m) = \mu(m)$.

Now, consider the matching $\mu \vee \mu^*$. We first show that $\mu \vee \mu^* \neq \mu$. If not, then $\mu \rhd_M \mu^*$. Since $\mu^* \in \pi(\succ)$ and $\mu^* \rhd_M \mu''$, this contradicts that μ'' is the $\rhd_M$-best matching in $\pi(\succ)$ such that $\mu \rhd_M \mu''$. Next, we show that $\mu \vee \mu^* \in \mathcal{S}(\succ^{\pi}) \setminus \pi(\succ)$. Recall that $\mu \in \mathcal{S}(\succ^{\pi})$ is given and we know that $\mu^* \in \pi(\succ)$. Then, as argued in the main text, $\mu^* \in \pi(\succ)$ implies that $\mu^* \in \mathcal{S}(\succ^\pi)$. Since $\mathcal{S}(\succ^\pi)$ is a lattice, it follows that $\mu \vee \mu^* \in \mathcal{S}(\succ^\pi)$. To see that $\mu \vee \mu^* \notin \pi(\succ)$, recall that $\mu' \rhd_M \mu \vee \mu^* \rhd_M \mu$. Moreover, $\mu' \neq \mu \vee \mu^*$, since $(\mu \vee \mu^*)(m) = \mu(m)$. Then, since μ' is the $\rhd_M$-worst matching in $\pi(\succ)$ such that $\mu' \rhd_M \mu$, we must have $\mu \vee \mu^* \notin \pi(\succ)$. Thus, we conclude that $\mu \vee \mu^* \in \mathcal{S}(\succ^\pi) \setminus \pi(\succ)$ such that $\mu \vee \mu^* \rhd_M \mu$ and $\mu \vee \mu^* \neq \mu$. But, this contradicts our choice of μ, thus we conclude that $\mu(m) \notin \pi_m(\succ)$.

Now, we are ready to show that $\mu \notin \mathcal{S}(\succ^\pi)$. Let $w = \underline{\mu}(m)$. Then, by the construction of $\succ^\pi$, m is the $\succ^\pi_w$-best man for w. Since $\mu(m) \notin \pi_m(\succ)$, we also have $w \succ^\pi_m \mu(m)$. Therefore, m and w form a blocking pair at μ.

7.3 Proof of Theorem 2

If part: To see that *equity undominance* is satisfied, let $\succ \in \mathcal{P}$ and let $\mu \in \mathcal{S}(\succ)$ that minimizes $-\sum_{i\in N} F_i(\mu(i))$. Since for each $i \in N$, F_i is unimodal with respect to $\succ_i$ with mode med_i^A, it follows that there is no $\mu' \in \mathcal{S}(\succ)$ such that for each $i \in N$, $\mu'(i)$ is closer to med_i^A than $\mu(i)$.

Only if part: Let $\succ \in \mathcal{P}$ be a problem, $\mathcal{S}$ and $\mathcal{R}$ be the associated sets of stable matchings and rotations.

Step 1. Let $\rho = [(m_1, w_1), (m_2, w_2) \dots, (m_k, w_k)]$ be a rotation. For each $i \in \{1, \dots, k\}$, define $\rho(m_i) = w_i$ and $s_\rho(m_i) = w_{i+1}$; $\rho(w_i) = m_i$ and $s_\rho(w_i) = m_{i-1}$, where the addition and subtraction in the subscripts is modulo k. Now, for each $i \in N$ and $\rho \in \mathcal{R}$, define

$$\phi_i(\rho) = \begin{cases} 0 & \text{if there is no pair with } i \text{ in } \rho, \\ -1 & \text{if } s_\rho(i) \text{ is closer to } med_i^A \text{ than } \rho(i), \\ 1 & \text{if } \rho(i) \text{ is closer to } med_i^A \text{ than } s_\rho(i). \end{cases} \tag{5}$$

Then, since each $\mu \in \mathcal{S}$ is obtainable from μ^M by eliminating the rotations in R_μ, we have for each $i \in N$,

$$|Rank_i^A(\mu(i)) - Rank_i^A(med_i^A)| = K + \sum_{\rho \in R_\mu} \phi_i(\rho) \tag{6}$$

where $K = |Rank_i^A(\mu^M(i)) - Rank_i^A(med_i^A)| = \lfloor |A_i|/2 \rfloor - 1$.

For each $i \in N$ and $j \in A_i$, let ρ_{ij} be the unique rotation elimination of which makes i matched to j. By construction of ϕ_i, for each $\rho \in R_\mu$, if there is no pair with i, then $\phi_i(\rho) = 0$ and if there is a pair with i, then $\rho \to \rho_{i\mu(i)}$. Therefore, we have

$$|Rank_i^A(\mu(i)) - Rank_i^A(med_i^A)| = K + \sum_{\{\rho | \rho \to \rho_{i\mu(i)}\}} \phi_i(\rho). \tag{7}$$

Step 2. By using the generality of our results in Section 7.1, we construct an additive function $G : Cl(\mathcal{R}) \to \mathbb{R}$ for π such that $G(\emptyset) = 0$ and for each nonempty $R \in Cl(\mathcal{R})$, $G(R) = \sum_{\rho \in R} g(\rho)$. To see this, let $\rho \in \mathcal{R}$ and N_ρ^+ (N_ρ^-) be the set of agents such that $\phi_i(\rho) > 0$ ($\phi_i(\rho) < 0$). First, note that for each $\rho \in \mathcal{R}$, if $N_\rho^+ = \emptyset$ ($N_\rho^- = \emptyset$), then we must have $g(\rho) < 0$ ($g(\rho) > 0$), otherwise some agent $i \in N_\rho^-$ ($i \in N_\rho^+$) must receive a positive (negative) $g_i(\rho)$ value, contradicting that the resulting F_i is unimodal. To guarantee that this is not the case, for each $\lambda \in \Lambda$ that is not singleton choose $\lambda^{\downdownarrows}$ as the union of $\lambda^{\downarrow}$ and all $\rho \in \lambda$ such that $N_\rho^+ = \emptyset$, and $\lambda^{\upuparrows}$ as the union of $\lambda^{\uparrow}$ and all $\rho \in \lambda$ such that $N_\rho^- = \emptyset$, i.e. $\lambda^{\downdownarrows} = \lambda^{\downarrow} \cup \{\rho' \in \lambda \mid N_{\rho'}^+ = \emptyset\}$ and $\lambda^{\upuparrows} = \lambda^{\uparrow} \cup \{\rho \in \lambda \mid N_{\rho'}^- = \emptyset\}$. Note that, we have $\lambda^{\upuparrows} \cap \lambda^{\downdownarrows} = \emptyset$, as we know that $\lambda^{\uparrow} \cap \lambda^{\downarrow} = \emptyset$. Now, by Lemma 10, for the associated mapping H, for each $\mu \in \mathcal{S}$ with $R_{\bar{\mu}} \subset R_\mu \subset R_{\underline{\mu}}$, we have $H(R_\mu) = 0$ if and only if $\mu \in \pi(\succ)$.

In moving to G from H, there is a minor problem of directly using the construction presented at the end of Section 7.1. To fix this, we modify our construction of g by using *equity undominance* to guarantee that each F_i can be constructed as to be unimodal. Let $\mu \in \pi(\succ)$ and $\rho \in \mathcal{R}$, first, we observe that: **(i)** If ρ is exposed in μ, then $N_\rho^+ \neq \emptyset$; otherwise let $\mu' = \mu \circlearrowleft \rho$, then for each agent $i \in N$ with $\mu(i) \neq \mu'(i)$ we have $\mu'(i)$ is closer to med_i^A than $\mu(i)$, which contradicts that π satisfies *equity undominance.* **(ii)** If μ is obtained from another matching μ' by eliminating ρ, i.e. $\mu = \mu' \circlearrowleft \rho$, then $N_\rho^- \neq \emptyset$; otherwise for each agent $i \in N$ with $\mu(i) \neq \mu'(i)$ we have $\mu'(i)$ is closer to med_i^A than $\mu(i)$, contradicting that π satisfies *equity undominance.*

Now, let $\rho \notin R_{\underline{\mu}}$ ($\rho \in R_{\bar{\mu}}$). Then, as it was defined in Section 7.1, $g(\rho) = 1$ ($g(\rho) = -1$) even if $N_\rho^+ = \emptyset$ ($N_\rho^- = \emptyset$). To fix this, first, update g such that if $N_\rho^+ = \emptyset$ ($N_\rho^- = \emptyset$), then $g(\rho) = -1$

$(g(\rho) = 1)$. However, following this update, it should still be the case that if $R \in Cl(\mathcal{R})$ minimizes G, then $R_{\bar{\mu}} \subset R \subset R_{\underline{\mu}}$. To guarantee this, if $\bar{\mu} \neq \mu^M$, then let $\rho \in \mathcal{R}$ be such that $\bar{\mu} = \mu' \circlearrowright \rho$ for some $\mu \in \mathcal{S}$. By (ii), we have $N_\rho^- \neq \emptyset$. Therefore, we can pick $g(\rho)$ small enough to guarantee that for each $R \in Cl(\mathcal{R})$, if $R \subsetneq R_{\bar{\mu}}$, then $G(R_{\bar{\mu}}) < G(R)$. Similarly, if $\underline{\mu} \neq \mu^W$, then let $\rho \in \mathcal{R}$ such that ρ is exposed in $\underline{\mu}$. By (i), we have $N_\rho^+ \neq \emptyset$. Therefore, we can pick $g(\rho)$ big enough to guarantee that for each $R \in Cl(\mathcal{R})$, if $R_{\underline{\mu}} \subsetneq R$, then $G(R_{\underline{\mu}}) < G(R)$.

Step 3. First, for the mapping $g : \mathcal{R} \to \mathbb{R}$ that is constructed in the previous step, we show that for each $\rho \in \mathcal{R}$, there exists $\{g_i(\rho)\}_{i \in N}$ that satisfies:

$$\sum_{i \in N} g_i(p) = g(\rho), \text{and} \tag{8}$$

$$\text{for each } i \in N,\ g_i(\rho) = \alpha\phi_i(\rho) \text{ for some } \alpha > 0. \tag{9}$$

To see this, recall that g is constructed such that for each $\rho \in \mathcal{R}$, if $N_\rho^+ = \emptyset$ ($N_\rho^- = \emptyset$), then $g(\rho) < 0$ ($g(\rho) > 0$). Now, for each $i \in N$ and $\rho \in \mathcal{R}$, define $g_i(\rho)$ such that if $g(\rho) < 0$, then

$$g_i(\rho) = \begin{cases} 0 & \text{if } \phi_i(\rho) = 0, \\ \dfrac{(g(\rho) - |N_\rho^+|)}{|N_\rho^-|} & \text{if } \phi_i(\rho) < 0, \\ 1 & \text{if } \phi_i(\rho) > 0. \end{cases} \tag{10}$$

If $g(\rho) \geq 0$, then

$$g_i(\rho) = \begin{cases} 0 & \text{if } \phi_i(\rho) = 0, \\ -1 & \text{if } \phi_i(\rho) < 0, \\ \dfrac{(g(\rho) + |N_\rho^-|)}{|N_\rho^+|} & \text{if } \phi_i(\rho) > 0. \end{cases} \tag{11}$$

Note that, by this construction, $\{g_i(\rho)\}_{i \in N}$ satisfies (8) and (9). Next, recall that for each $i \in N$ and $j \in A_i$, ρ_{ij} was the unique rotation elimination of which makes i matched to j. Then, we define

$$F_i(j) = \sum_{\{\rho | \rho \to \rho_{ij}\}} g_i(\rho). \tag{12}$$

It directly follows from (7) and (9) that for each $i \in N$, F_i is unimodal with mode med_i^A. Finally, let $\mu \in \mathcal{S}$ and define $F(\mu) = \sum_{i \in N} F_i(\mu(i))$. To see that $F(\mu) = G(R_\mu)$, first, note that since $R_\mu \in Cl(\mathcal{R})$, we have $R_\mu = \bigcup_{i \in N} \{\rho | \rho \to \rho_{i\mu(i)}\}$. By (9), for each $i \in N$ and $\rho \in \mathcal{R}$, if $\phi_i(\rho) = 0$, then $g_i(\rho) = 0$. Therefore, it follows from (8) that

$$G(R_\mu) = \sum_{\rho \in R_\mu} g(\rho) = \sum_{i \in N} \sum_{\{\rho | \rho \to \rho_{i\mu(i)}\}} g_i(\rho). \tag{13}$$

By substituting (12) into (13), we conclude that $F(\mu) = G(R_\mu)$.

7.4 An example

We present a problem to show that several claims made throughout the main text hold. Consider the problem with eight men and women whose preferences are represented by the table in Figure 9, where each entry is associated with a man m and a woman w. If m and w are attainable for each other, then the rank of w in $\succ_m$ (the rank of m in $\succ_w$) is written in the bottom (top) corner. If m and w are unattainable, then the associated cells are shaded, indicating that we can freely choose the associated rank as far as it is bigger than the number of agents' total attainable mates. Note that each agent has a unique median attainable mate in this problem, the associated median attainable ranks are boxed in the table.

M \ W	a	b	c	d	w	x	y	z
1	1 \ 3	2 \ 5			4 \ 2		$\boxed{3}$ \ $\boxed{4}$	5 \ 1
2	4 \ $\boxed{2}$	$\boxed{3}$ \ 6			2 \ 4	5 \ 1		1 \ 5
3		$\boxed{3}$ \ $\boxed{4}$			5 \ 1	1 \ 3	2 \ 5	4 \ 2
4	5 \ 1				1 \ 5	4 \ $\boxed{2}$	$\boxed{3}$ \ 6	2 \ 4
5		$\boxed{3}$ \ 1	1 \ 3				2 \ 2	
6		2 \ 2		1 \ 3			$\boxed{3}$ \ 1	
7		1 \ 7	5 \ 1	4 \ $\boxed{2}$			$\boxed{3}$ \ 3	2 \ $\boxed{3}$
8		$\boxed{3}$ \ 3	4 \ $\boxed{2}$	5 \ 1	2 \ $\boxed{3}$		1 \ 7	

Figure 9: The problem.

In Figure 10, we present the associated rotations and their poset. For each agent i that appears in a rotation, the superscript $(+)$ ($(-)$) means that i gets far away (closer to) from their attainable median. For example, $(m^+, w^-) \in \rho$ means that m gets far away from his attainable median, whereas w gets closer to her attainable median after the elimination of rotation ρ.

$$\rho^1 = [(2^{(-)}, z^{(-)}), (4^{(-)}, w^{(-)})]$$

$$\rho^2 = [(2^{(-)}, w^{(-)}), (7^{(-)}, b^{(-)}), (4^{(-)}, z^{(-)}), (8^{(-)}, y^{(-)})]$$

$$\rho^{31} = [(1^{(-)}, a^{(-)}), (2^{(+)}, b^{(-)})];\ \rho^{32} = [(3^{(-)}, x^{(-)}), (4^{(+)}, y^{(-)})]$$

$$\rho^{41} = [(1^{(-)}, b^{(-)}), (3^{(-)}, y^{(-)})];\ \rho^{42} = [(2^{(+)}, a^{(+)}), (4^{(+)}, x^{(+)})]$$

$$\rho^5 = [(1^{(+)}, y^{(+)}), (8^{(-)}, w^{(+)}), (3^{(+)}, b^{(+)}), (7^{(-)}, z^{(+)})]$$

$$\rho^{61} = [(1^{(+)}, w^{(+)}), (3^{(+)}, z^{(+)})];\ \rho^{62} = [(8^{(+)}, b^{(+)}), (5^{(-)}, c^{(-)}), (7^{(+)}, y^{(+)}), (6^{(-)}, d^{(-)})]$$

$$\rho^{71} = [(8^{(+)}, c^{(+)}), (7^{(+)}, d^{(+)})];\ \rho^{72} = [(6^{(+)}, b^{(+)}), (5^{(+)}, y^{(+)})]$$

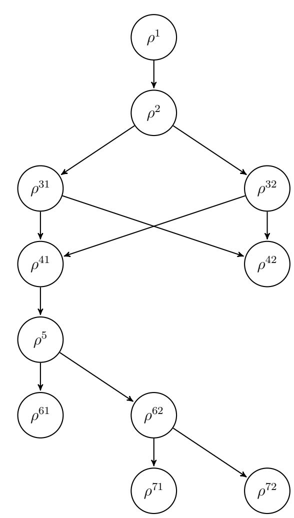


Figure 10: The rotation poset.

In Figure 11, we present the stable matching lattice associated with the problem in Figure 9 such that each stable matching is represented as an array $[w_1, \ldots, w_8]$, where each w_i is the woman who is matched with man i. Each edge is labeled by the associated rotation whose elimination from the matching in the upper end of the edge results in the matching in the lower end of the edge. The green (lighter) colored matchings are the equity-undominated ones. For this problem, we make the following observations.

1. The set of equity-undominated matchings is not a sublattice of the original problem (claimed in Footnote 14). The matching $[bayxcdzw]$ is the meet of two equity-undominated matchings. However, it is equity dominated by $[yabxcdzw]$. It also follows from this observation that a stable matching that is between two equity undominated matching, according to the men-wise better than relation, can be equity dominated.
2. The unique stable matching that is chosen by the equal weight median rule minimizing the total distance from the median is $[yabxcdzw]$.
3. The stable matching rule presented in Example 6, which chooses the set of attainable sex-equal stable matchings, does not satisfy *equity undominance* (claimed in Footnote 13). The unique attainable sex-equal stable matching is $\mu^* = [yxbacdzw]$, since $\sum_{m \in M} Rank_m^A(\mu^*(m)) = \sum_{w \in W} Rank_w^A(\mu^*(w)) = 22$. However, μ^* is equity dominated by the matching $[bxyacdzw]$.
4. Every mixture of stable matchings is not necessarily stable (claimed in Footnote 11). To show this, we modify the problem so that the rank of woman d for man 5 is 2 and the rank of man 5 for woman d is 3. We claim that the stable matching lattice remains unchanged after this modification. We show this by showing that this modification has no effect on the set of rotations. To see this, first note that ρ^{62} the first rotation that contains 5, and it also turns out to be the first one that contains d. Now, note that once ρ^{62} is eliminated, 5 is matched to y, where $d \succ_5 y$, and d is matched to 7, where $7 \succ_d 5$. It follows that there can be no rotation that contains $(5, d)$. Put differently, 5 and d remain unattainable for each other after the modification, and thus the rotation poset remains unchanged. Next, consider the stable matchings $\mu' = [yabxcdzw]$ and $\mu'' = [wazxybdc]$. Then let μ be the matching $[wabxydzc]$ that is obtained as a mixture of μ' and μ'', in the sense that for each agent i, we have $\mu(i) \in \{\mu'(i), \mu''(i)\}$. Clearly, μ is not stable, since it is not stable in the original problem. Alternatively, to directly see that μ is not stable, note that $(5, d)$ forms a blocking pair in μ.

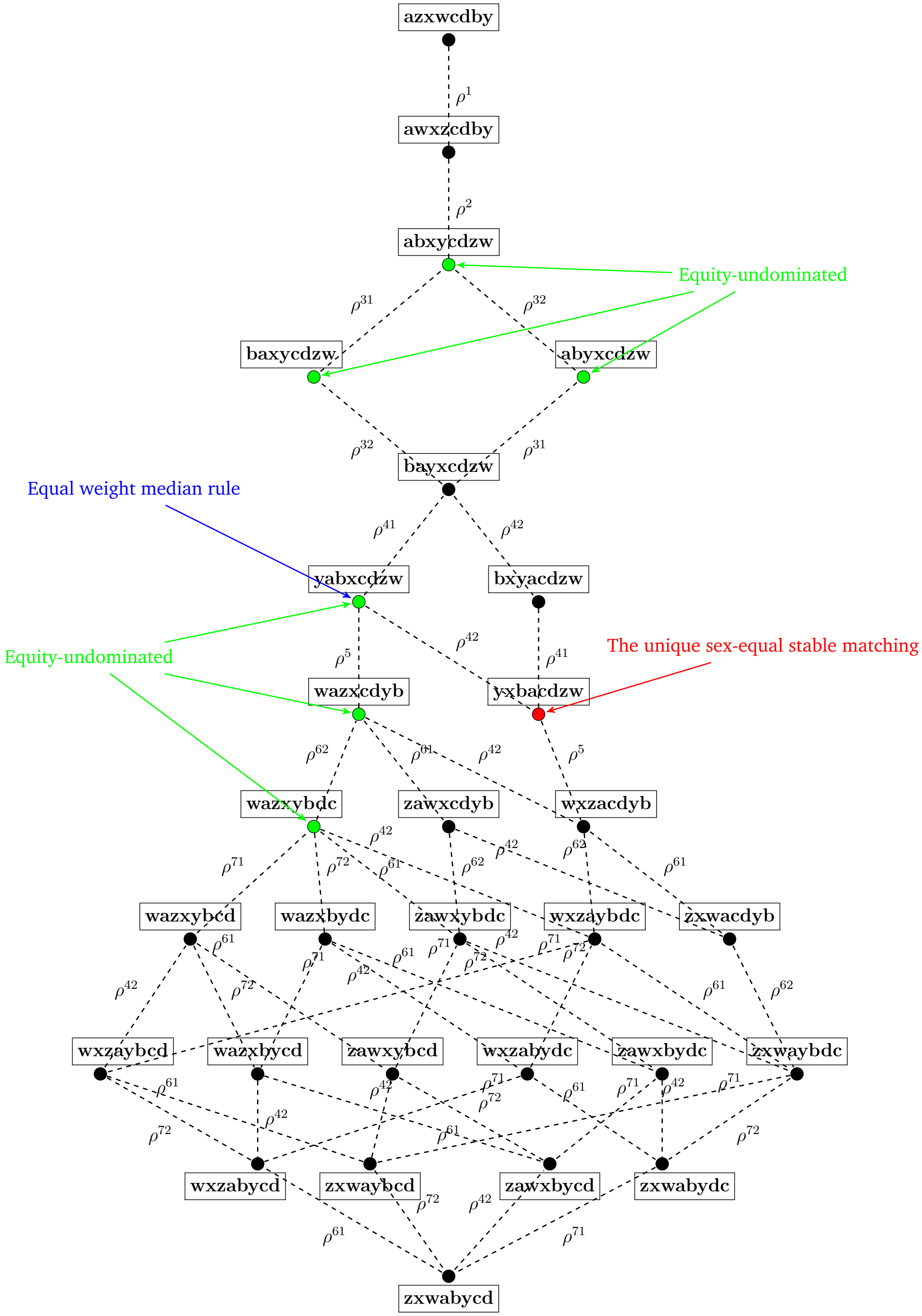


Figure 11: The associated stable matchings lattice.